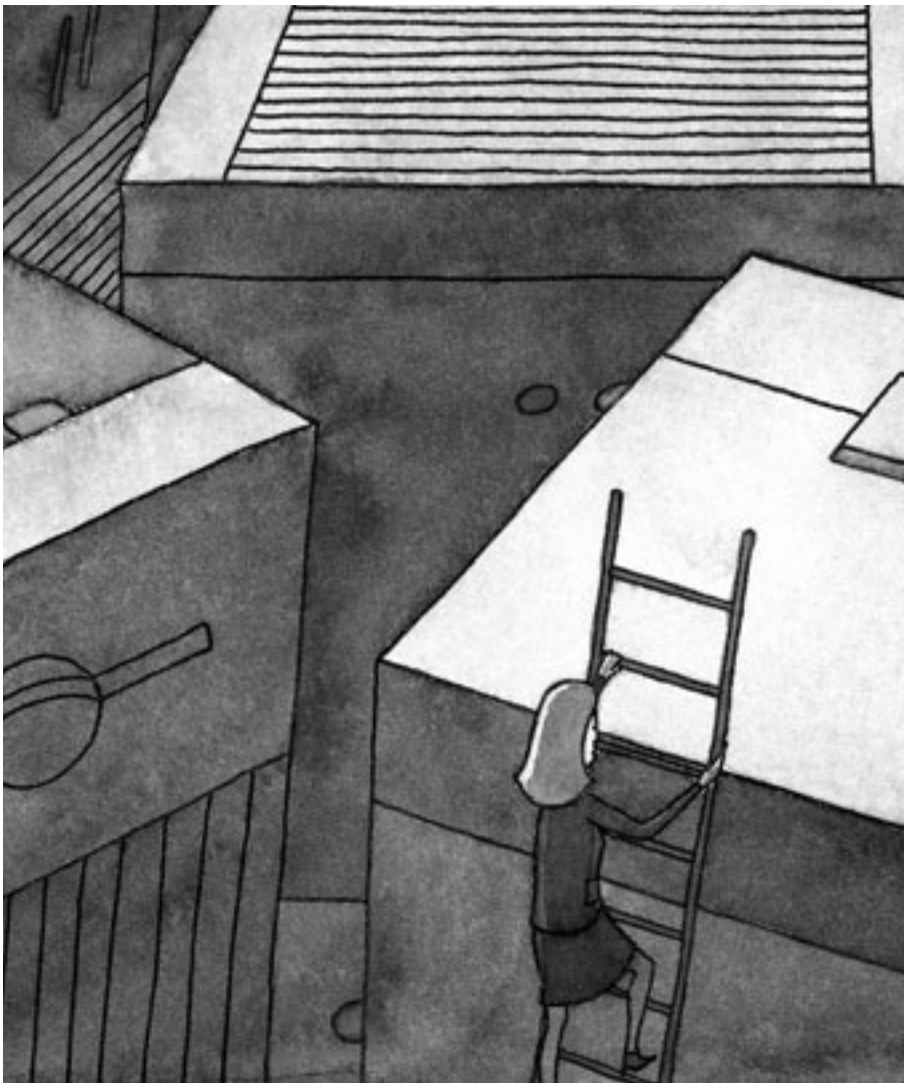
What Academia Can Gain from Building a **Data Warehouse**

Used effectively, data warehouses can be a significant component of strategic decision making on campus

By **David Wierschem, Jeremy McMillen,** and **Randy McBroom**



Organizations exist in a dynamic environment that has experienced significant changes in the past 10 years. Advances in technology have made the world smaller, competition has increased, the pace of change has quickened, accountability continues to rise, and customers are becoming more demanding and less patient. These changes, as well as others, are forcing organizations to identify and adopt new strategies for growth and survival.

One of the most important components to facilitate the development of these strategies is information. Information is required to identify where the organization has been, where it is now, and where it needs or wants to be in the future.

The business community has been investigating the use of data warehouses for some time in an attempt to glean insight into customer behavior, analyze supply chain activities, and support performance measurement systems, as well as other business process operations.[1] Since their initial development in the 1970s,[2] data warehouses have become a significant component of strategic decision making for business.

Major users of data warehouses include credit card companies, retailers, financial services, banks, airlines,

manufacturers, telephone companies, and insurance companies.[3] Markedly deficient in the literature's list of high-profile users are academic institutions. While several universities have developed or are currently developing a data warehouse,[4] they are primarily major research universities. One reason for the absence of data warehouses in the academy may be the perceived insulation academic institutions have from the competitive pressures of business. Another reason may be the attitude of academic administrators who view institutions of higher education as above the common demands of business. However, in evaluating the purpose and applications of these major users, it appears that academic institutions share similar, if not identical, informational demands.

In this article we identify the opportunities associated with developing a data warehouse in an academic environment. We begin by explaining what a data warehouse is and what its informational contents may include, relative to the academic environment. Next, we address the current environmental drivers that provide the opportunities for taking advantage of a data warehouse and some of the obstacles inhibiting the development of an academic data warehouse. Finally, we provide strategies to justify developing a data warehouse for an academic institution.

Definition of a Data Warehouse

In its simplest form, according to Gallegos,[5] "A data warehouse is a collection of integrated databases designed to support managerial decision making and problem solving functions." In application, data warehouses are much more than data storage; several specific attributes unique to data warehouses separate them from other data repositories. Noteworthy is the fact that data-warehouse information is generally static in nature. That is, unlike operational transaction processing databases that constantly change as new business transactions occur, data warehouses provide a stable snapshot in time of the collected data and information.

The primary purpose of the data warehouse is to support managerial decision making.

Another unique aspect of data warehouses is that the contents continue to expand as information is added to the existing data. Most transactional processing systems have a predefined cutoff time period for which active data are maintained; once that time period is reached, the aged data are archived. The expansive aspect of the data warehouse results in large and ever-growing amounts of data. Additionally, unlike traditional data repositories that focus on a specific function or application, data warehouses collect informational databases from throughout the organization, which allows for an integrated analysis of information from across all areas of the organization.

Finally, the primary purpose of the data warehouse is to support managerial decision making. While other databases are designed to facilitate transactions or to provide customer service, data warehouses are designed to support analysis of business and customer activities.

Thus a functional and useful data warehouse becomes a rather expansive repository of information. The operations associated with an academic organization generate large amounts of data and provide the ideal setting to support the development of a data warehouse. Transactional and operational academic databases providing input for an academic data warehouse could include the following:

- Admissions data
- Enrollment data
- Employment data
- Course data
- Financial aid data
- Housing data
- Alumni data
- Facilities data
- Student services data

Most academic institutions have databases in place that capture and record the associated individual transactions for each of these areas. The integration of these different areas, however, provides many opportunities to gain insight into the interactions between them.

Drivers for Developing an Academic Data Warehouse

As stated earlier, businesses have been using data warehouses since the 1970s. Academic institutions, however, only recently have begun to identify and explore the possibilities and benefits data warehouses offer. Several environmental factors have produced a series of driving forces that encourage academic institutional leaders to investigate data-warehousing options. These factors include decreases in governmental financial support, faculty supplies, and research funding, and increases in student tuitions, competition, faculty salaries, faculty support, and expectations from students, parents, and employers.

These factors have led academic administrators to change how they view higher education operations. As administrators strive to cope with these environmental factors, they must evaluate their informational needs relative to these factors and to accessible information. Each factor generates informational drivers. Drivers may be generated from a single factor or may result from several factors working together. However they are generated, the drivers support the need for the development of a data warehouse.

Change

Change is one of the most prominent drivers. For the academic community, change comes in many different forms and from many different sources, including federal and state regulations, student interest and preparedness, faculty development, industry support, employer skill demands, instruction materials and techniques, and financial support. Each source of change is influenced by different environmental factors, which in turn are changing. The result is a constant and complex environment of change.

While organizations in various sectors have adapted to this environment of change in differing degrees, academic

institutions have been slow to change. Historically, the structure of higher education's mission was such that competition between institutions was minimal, financial support was stable, and other environmental factors had little influence. Academic institutions only recently have been impacted by environmental forces generally affecting other sectors—they are now being held accountable to the same degree as nonacademic institutions.

For example, public institutions of higher education are witnessing continuing decreases in governmental funding support with concomitant increases in tuition. Additionally, universities must now compete with industry to attract qualified faculty and staff, especially in the areas of technology and science. State and federal funding agencies are being held more accountable for the tax dollars they spend and are increasingly demanding evidence of improvement from the institutions they fund. Student demographics continue to fluctuate, reflecting changes in the structure of the family and demographic shifts.

Finally, technology has dramatically altered the academic curricula, affecting both what and how content is taught. New courses addressing technology, such as HTML, e-commerce, and mobile communications, as well as the enhancement of classic subjects such as physics and chemistry, have required increased investments in funding and time to incorporate new information into the classroom. Additionally, the Internet, classroom projection systems, laptop computing, and distance education are changing how and where classes are taught.

The majority of changes affecting higher education have come to light only recently, and it appears that the pace of change will continue to accelerate. This change acceleration is a driving force for the gathering of information to support strategies and processes that address change. Immense amounts of data are now generated, both internally and externally. Integrating this information into administrative planning and decision making is more important now than ever. An academic data warehouse provides the perfect tool for capturing, organizing, and analyzing this information for decision makers.

Control

Control is one of the most controversial issues associated with organizational management, especially for the academic community. Individual faculty members, department heads, deans, and administrators believe it is important to control the forces that affect their success. For faculty members, control is the ability to design the content and structure of classes and to conduct research in the ways they deem best. For department heads, control is the ability to direct faculty assignments for courses, curricular and program assessment and change, the structure and exceptions for student degree plans, and the allocation of departmental resources. For deans, administrators, and administrative staff, control is the ability to manage resources according to their desires. Decisions made to benefit an individual unit, however, often fail to result in the optimal use of resources for the organization as a whole.[6,7]

To address this issue, a shift in the decision-making structure, often to a centralized or single point of decision making, is needed to facilitate global or organization-wide resource allocation and utilization. This change in control structure requires more information. Previously, individual decision makers had intimate knowledge of situations and were able to make decisions quickly and authoritatively. Centralized decision making, however, requires effort to gather and organize information for decision makers who are not knowledgeable of specific areas.

A data warehouse provides a centralized and standardized data repository that meets the needs of the centralized decision maker. The warehouse also offers a structured data model across the entire organization, providing the decision maker with a consistent and global view of the organization.

Planning

Increasingly, academic institutions are being asked to define their mission, state their purpose, and plan their future. Accreditation bodies, alumni groups, and advisory boards are requiring universities to develop strategic plans that provide direction for decision making. Academic strategic plans typically cover all areas of the organization, including the libraries, colleges, departments, programs, services, and budget. Moreover, the importance of planning activities will likely continue to increase.

The broad reach of strategic plans requires large amounts of data to identify and support their many specific components, including enrollments, retention, space utilization, growth, cooperative agreements, and student outcomes assessment. For each component, historical data are used to explain the past trends, current position, and projections.

Gathering the data necessary to generate these plans and to be used for other purposes is an exceptionally time-consuming and labor-intensive activity. A data warehouse can make gathering and organizing the vast amounts of data required for planning easier and faster. The data warehouse can be structured specifically to support the gathering, storage, and analysis of the large amounts of historical data often needed for strategic planning activities.

Accountability

Accountability, to some degree, has always been a part of modern academic life. Grades, the tenure process, and merit pay are all accountability measures that have been institutionalized. Additionally, new demands for accountability are being placed on the academic community. Accreditation bodies are requiring increasingly specific documentation detailing faculty and student activities. Federal and state governments are imposing more stringent guidelines and demanding more and better results for the tax-dollar investment in public higher education. Likewise, governing bodies and alumni are holding private institutions more accountable.

Even in public higher education, stakeholders such as alumni groups, advisory boards, and other supporters are requiring measured improvements or

successes in exchange for their participation. These increased measures require large amounts of data to be collected, stored, and organized in such a way that timely and accurate reports can be generated for each of the growing number of interested parties. Structured and administered properly, an academic data warehouse can provide both the consistency of stable data and the security control required to address accountability issues.

Decision Making

By far the most pervasive driver for developing an academic data warehouse is the necessity for continuous decision making and quality enhancement. The ability to perform analysis on historical data to make better decisions, to improve productivity, and to enhance operational effectiveness is required for today's dynamic organizations. Academic institutions are facing new forces (for example, distance education and Web-based instruction) that subject them to the environmental factors of the competitive marketplace.

Current business practices are predicated on fast decision making supported by accurate and comprehensive data. Additionally, flexibility to rapidly respond to the fast-changing dynamics of the environment requires the ability to quickly search for and identify new opportunities. Corporate databases are full of historical information containing insight into customer behavior and potential opportunities. Likewise, academic histories hold many opportunities for improving effectiveness through better and more responsive decision making. A data warehouse supports this analysis by storing the vast histories of institutions, as well as the volumes of external data also necessary for evaluation and decision making.

The drivers listed above—change, control, planning, accountability, and decision making—are not all-inclusive. There are many others not specifically covered, such as benchmarking and needs assessment, that can be categorized under one of the above. These key drivers affect all institutions of higher education. They work together to highlight the need for timely, structured, reliable data. In an information-intensive climate, success is often determined by what and how data are used. Given this, the academic data warehouse is one of the few tools available that can support the importance and increasing volume of data that decision makers require today and in the future.

Obstacles to Developing an Academic Data Warehouse

Accepting the precepts of the above drivers then begs the question of why academic institutions are not employing data warehouses. Five primary factors explain why academic institutions fail to take advantage of data warehousing.

Fear of Change

Perhaps the biggest obstacle for academic institutions is the fear of change. In an environment where processes and procedures have remained the same for decades, the prospect of implementing new ones can be intimidating. In data-intensive areas such as registration, admissions, and financial aid, any change to the type or form of the data can increase the workload considerably. In many instances, current processes are followed simply by force of habit. Any proposed change will be assessed relative to the time it will take to learn the new process or procedure, how it will affect the daily workload, and what purpose it will serve.

Including those who will be affected by proposed changes in the planning process can reduce, if not eliminate, the apprehension level. At the very least, it is important that the lines of communication stay open and that information about progress is communicated. Failure to address these concerns can result in sabotage by those responsible for implementation.

Control and Ownership

The value of the information is another major issue. Typically, whoever owns or controls the data has an intrinsic attachment to or responsibility for it, which often leads to a proprietary view of the data. Consequently, any attempt by others to use or manipulate the data is perceived as a threat. In today's information-intensive environment, where "information is power," the perception is that those who possess the information have the power. Any situation that results in the easy access of information threatens those currently in power. Therefore, for a data warehouse to be effective, a culture that values cooperative sharing of information is required.

Building a cooperative culture takes time and requires the development of trust among involved parties. A key component in this process is to identify a champion, or top management individual, who supports and drives the warehouse development process. It is also important to communicate clearly and effectively the purposes of the data warehouse, reassuring those currently in control of the data. Finally, incorporating all involved parties in the design and development of the data warehouse will facilitate the cooperative culture needed for success.

Fear of Accountability

Not as obvious, nor as politically correct to discuss, is the fear of accountability. Frequently, the purpose of collecting data is for measurement or evaluation. Data may be used for evaluating programs, personnel, or identified outcomes as defined by the institution's administration. When an individual is approached to obtain access to a data set, a common response is to question why the information is being requested. The response will be viewed with suspicion if the information is needed for anything other than meeting a specific, identifiable output. Often, fear that the data will be used to further evaluate an individual's job or effectiveness results in apprehension in providing the requested data.

The ability and extent to which this issue can be addressed depends on the purpose of the data warehouse. If the purpose is in fact to impose new accountability measures, then a period of adjustment needs to be clearly identified and communicated. Specific measures and criteria must be delineated so that those affected have time to adjust. If the purpose is not for accountability,

then the actual purpose of the data warehouse should be clearly communicated.

Cost

While the previous issues may cause apprehension in developing an academic data warehouse, cost factors significantly in the failure to continue development. It can even lead to elimination of a data warehouse. A review of business journals suggests that a data warehouse requires millions of dollars to develop, plus significant hardware and personnel investment, in order to maintain and use it. The current fiscal restrictions in the academic community make such an undertaking extremely difficult, if not impossible.

There are many ways to address the cost issue, however, depending on the level of interest and available resources. The overall scope of the project can be reduced so that it is financially feasible yet still useful. The project also can be broken down into smaller components and developed over a longer period of time. Other possibilities are contingent on the resources available, time frame desired, and scope.

Complexity

Finally, the enormity and complexity of designing and developing a data warehouse can be intimidating. This is primarily due to the broad range of sources from which data will come, including admissions, the registrar's office, financial aid offices, state offices, federal offices, high schools, colleges and departments, alumni, job placement, and many more. Additionally, because of the evolutionary design of educational information systems and academic institutions' independent organizational structure, many different systems will be used. This will result in many different data formats that must be analyzed, formatted, cleaned, and structured, including flat files, text files, spreadsheets, and many others. The independent systems that have been employed will require extensive analysis to understand and identify the many different labeling schemas.

While the innate nature of a data warehouse is complex, the use of a comprehensive and well-thought-out methodology applied to its design, development, and implementation can help immensely. In addition, a high degree of documentation tracking and support for all decisions and procedures will serve to make the project manageable and somewhat amenable to personnel changes.

The development of a data warehouse provides a centralized source of information.

Justification for Developing an Academic Data Warehouse

We have listed just a few of the obstacles facing academic administrations when considering building a data warehouse. While they may seem daunting, the benefits to be gained may easily outweigh the disadvantages.

Improved Customer Service

While not typically associated with academic institutions, customer service is a major component of the education process. In terms of academia, customer service would include, but is not limited to, student services, admissions procedures, and alumni, faculty, and staff support. The development of a data warehouse provides a centralized source of information and static formalized reporting processes that can facilitate rapid accumulation and dissemination of data. Examples would include support of degree plan audits, student transcripts, and class schedules, as well as annual reporting documents for governmental and other reporting organizations. Using a centralized data warehouse that is accessible across different academic units would provide faster responses and the ability to quickly analyze problems and develop solutions to meet the needs of the institution's various customers.

Strategic Planning

Supplying the data necessary for developing the institution's strategic plans is a key aspect of a data warehouse. Strategic plans are an increasingly important feature of academic administration. Accreditation organizations and advisory boards are focusing more on how organizations envision the future and prepare to meet the challenges they face.

Academic strategic planning must take into consideration the past and current states of the institution. A data warehouse provides a powerful tool to evaluate past decision making and to gain insight into the potential options and consequences of current decisions. Additionally, as new responsibilities and mandates are added, historical data provide valuable benchmarks for evaluating successes.

Improved Administration

The data warehouse's contribution to the everyday administration of the institution is significant. Providing quick access to historical data, as well as to the different data stores, enables administrators to make better decisions in a timely manner. This becomes more important as environmental pressures on the institution intensify. Such pressures as governmental mandates, new technology, increased employer involvement, and decreased funding all require timely and authoritative decision making.

A data warehouse, with its standardized and powerful analytical capabilities, provides an opportunity to quickly and reliably prepare for a proactive approach to decision making. External bodies, such as legislatures, investigating boards, and accreditation agencies, can be provided with required metrics in a quick and efficient manner. Additional data to enhance or augment specific issues or concerns can also be obtained quickly to better support the institution's position.

Reduced Costs

Finally, when a complete cost-benefit analysis is done on the development and implementation of an academic data warehouse, it is apparent that the benefits far outweigh the cost. Cost may in fact become the major driving force behind the development of a data ware-

house. For the University of Washington, for example, a budget reduction was the impetus for development of its academic data warehouse.[8] The majority of the costs are up front, associated with the initial design and development of the warehouse. Once the warehouse has been implemented, the costs to maintain it can be minimal, depending on the complexity of the different data sources and the amount of data included.

No single methodology best identifies the appropriate depth and breadth of data for a particular institution. With appropriate up-front design, however, the costs can be minimized.

Beneficiaries

The primary beneficiaries of the proposed academic data warehouse are the key decision makers of the institution. Administrators will benefit from improved access to and analytical capabilities of historical and environmental data. Strategic planning teams, along with operational decision makers, will be more informed and able to guide the institution more effectively into the future.

Depending on the design and purpose of the data warehouse, the majority of the faculty, staff, and students may not have direct access to the data. They will, however, benefit from the results of improved administrative decision making and planning. *e*

Endnotes

1. H.J. Watson et al., "Current Practices in Data Warehousing," *Information Systems Management*, Vol. 18, No. 1, Winter 2001, pp. 47–55.
2. H.J. Watson and B.J. Haley, "Managerial Considerations," *Communications of the ACM*, Vol. 41, No. 9, Sept. 1998, pp. 32–37.
3. L. Chen, T. Sakaguchi, and M. Frolick, "Data Mining Methods, Applications, and Tools," *Information Systems Management*, Vol. 17, No. 1, Winter 2000, pp. 65–70.
4. "Case Study: Stanford University Gains Higher Learning with a Data Warehouse to Manage Old Information," *InfoWorld*, Vol. 17, No. 31, July 31, 1995.
5. F. Gallegos, "Data Warehousing: A Strategic Approach," *Information Strategy: The Executive's Journal*, Vol. 16, No. 1, Fall 1999, pp. 41–47
6. M.A. Khan and A. Rustichini, "Cournot-Nash Equilibrium Distributions of Games with Uncertainty and Imperfect Information," *Journal of Mathematical Economics*, Vol. 22, No. 1, 1993, pp. 35–59.
7. H. Konishi, B.M. Le, and S. Weber, "Equilibria in a Model with Partial Rivalry," *Journal of Economic Theory*, Vol. 72, No. 1, Jan. 1997, pp. 225–237.
8. D. Friedman and P. Hofeman, "The Politics of Information," *Change*, May/June 2000, pp. 51–57.

David Wierschem (david_wierschem@tamu-commerce.edu) is an assistant professor in the Department of General Business and Management Information Systems at Texas A&M University–Commerce. Jeremy McMillen is Director of the Office of Instruction Technology, and Randy McBroom is Assistant to the President, Planning and Institutional Effectiveness, at Texas A&M University–Commerce.