Effective Practice Detail

Title:	**The George Washington University Student Data Mart**
Institution:	The George Washington University
Date Submitted:	2/15/2002
Subject Terms:	Data Administration and Management , Data Warehouse , Decision Support Systems , Student Information Systems
Background/Challenge:	Like most higher education institutions, The George Washington University (GW) is facing a new set of challenges that are more complex than ever. Overcoming fierce competition for the best students, meeting diverse student needs, finding adequate funds to meet constantly shifting demands, and increased accountability from internal and external constituents are affecting all aspects of the GW's business processes from recruitment to financial management. The complexity of these challenges requires continuous improvement of operational strategies based on accurate and timely decisions. In order to create a high-caliber decision support environment, GW is turning to data warehousing technology to deliver key performance metrics to end users on-demand and at their desktops. GW faced several challenges in the process: 1. GW's historical data are buried in heterogeneous data sources that support transactional day-to-day business. Data are difficult to extract and even more difficult to integrate into a single nomenclature due to the inconsistencies inherent in distributed data storage and ownership. 2. Transactional applications do not store data in models that support on-demand and ad hoc aggregations that are critical for generating key performance metrics. 3. Programmers or specialized reporting analysts in various GW departments are often asked to create reports using proprietary programs or highly technical applications. This process typically takes days if not weeks, and may produce compromised results due to inconsistent application of business rules or incomplete understanding of the relationships in the data. This process is costly to GW in terms of time and resources involved and ultimately does not serve the needs of the end users effectively. 4. Systems groups at GW, like at other institutions, are constrained by limited funds and their budgets are further strained by increasing demand for reports.
Practice/Solution:	Overview GW's solution for meeting these challenges is to create an enterprise data warehouse over time and in an incremental manner. The first phase of our solution, called the Student Data Mart (SDM), was rolled out to the user community in July 2001. The SDM uses data warehousing technology to: * Integrate raw data into an unified data model to support a set of key academic metrics * Create a user-friendly reporting environment for ad hoc analysis of data. * Deliver information to end users on-demand, through the GW Intranet, in commonly used formats such as Acrobat PDF, text files, and Excel spreadsheets. The objective was to provide clean and integrated data that can be accessed by business users in a business context, requiring no knowledge of databases, programming languages, or other technical know-how. An added objective was to create an enterprise-wide solution in increments and spread the investment over time, while relieving the pressure on the systems staff and providing end users direct access to the key performance metrics. The GW SDM user community has expanded significantly in the first 6 months after rollout. Users have consistently provided positive feedback and accolades on the success and robustness of the application. The Executive Director of Administrative Applications sponsored the project. Users from Institutional Research, Registrar's Office, Graduate Admissions, and individual colleges contributed significantly to the success of the project. The SDM vision-to-rollout duration was 18 months. Scope

The SDM contains ten years of recruitment, admissions, enrollment, registration, and GPA information for all students, all campuses, and all programs. It supports a wide-range of academic metrics, including per-campus and unduplicated enrollment counts, admissions selectivity, course enrollment data, student achievement and program growth. These metrics are directly related to academic goals of departments and of the university as a whole. The summary data can be "sliced and diced" across a variety of dimensions such as campus, college, level, degree, major, demographics, inquiry source, academic standing, student type, course/section, schedule type, and registration status among others. The SDM includes several multidimensional views for trend analysis of historical data.

Also, the SDM includes comprehensive "metadata"--information about the data contents--including business definitions of data elements and, for more technical audiences, "what's this element" mappings of SDM contents to their source elements in the GW transactional application.

Secure Single Point of Entry

Users access features of the SDM through Cognos' Upfront web portal product. This entry point provides links to standard reports that can be executed on current data, multidimensional views of summarized data, and end-user help such as frequently asked questions, quick reference guides, training material, what's-this-field crosswalks, and other useful links. Access to subject areas, views, reports, and individual data elements is controlled by a business-policy-driven security matrix, which authorizes access to data based on assigned user classes and authorizations that permit access through the firewall.

SDM Technologies

Reports and views are accessed by users through Web-based, zero-footprint business intelligence products from Cognos Corporation. This paradigm centralizes information dissemination to a server-based architecture. The need for software installation on individual desktops across the university campus is limited to a small set of "power" users who need to create specialized reports for other users. This approach affords GW a cost-effective way of rolling out the SDM user interface to a growing user community.

SDM data are stored in dimensional models to support high-volume, high-performance queries in an Oracle 8i relational database. Informatica Corporation's PowerMart 5.1 product suite is used to extract, transform, and load data from the source application to the warehouse. Business intelligence components are developed in Cognos Corporation's version 6 Web-based product suite (Upfront, Impromptu Web Reports, PowerPlay). The SDM is hosted on a distributed architecture using a multi-processor Sun E5500 server with Solaris operating system and two Windows NT servers.

Benefits:

The SDM offers several benefits:

1. By integrating data into a clean repository and by disseminating information over the Intranet, the SDM allows users to directly access data to measure key academic metrics on-demand at their desktops. Based on these metrics, users are able to make decisions with more precision and in a timely manner. By promoting a flexible reporting and analytical environment, GW ultimately hopes to relieve users of tedious processes of requesting reports.

2. SDM data integration processes and reports have exposed many data quality issues that are a natural outcome of distributed data ownership. Without the SDM it would be difficult to investigate the nature and extent of the data quality issues. With the SDM, GW's data quality staff is able to identify these issues more readily and that in turn is enabling the staff to define more accurate enterprise-wide data quality standards and procedures.

3. The SDM promotes the concept of reusable reports which allows users to share and leverage useful reports with ease and without programmer or technical intervention. For example, if an SDM user perceives that a newly created report can benefit other users, the user can easily publish the report for public use to the SDM portal. Reusable reports are allowing GW to shift focus from the creation of "one-off" reports that are costly to maintain to a cost-effective publish-and-subscribe paradigm. In addition to reports, multidimensional views in the SDM offer an exploratory environment that executives are using to identify trends and mitigate risk in time.

4. By July 2002, GW expects to be able to eliminate up to one third of all

	modifications and customizationsof the purchased Banner application that produce reports of student data. Eliminating the need to maintain these modifications/ customizations represents a significant cost savings to the software development organization. 5. In the future, the use of data warehouse technology will allow GW to tie student, accounts receivable, human resources, financial, asset management, and other data domains across common business attributes to create an enterprise-wide data warehouse. This type of integration is difficult, if not impossible, to achieve with the disparate source applications, but it is essential for decision-making.
References - URLs:	http://www.gwu.edu/~aapp/sdm
Costs:	The project team has secured $2 million in funding over the last 2 years. Future plans include a five-year budget projection and the rollout of additional data marts and a significant expansion of the user base. The project team consists of 1 full-time project manager and 5 project team members.
Replicable:	How replicable is this practice or solution? ○ 1 ○ 2 ○ 3 ◉ 4 ○ 5 **Not at all** **Highly**
Effectiveness:	How effective is this practice or solution? ○ 1 ○ 2 ◉ 3 ○ 4 ○ 5 **Not at all** **Highly**
Notes:	The GW SDM project is a model for other higher education institutions on three fronts: project management, project approach, and technical capital. Project Management Best practices related to project management, lessons learned, and critical success factors from our solution would be of value to other institutions embarking on similar projects since most higher education institutions probably face shortage of sponsorship, funds, and resources for data warehouse projects as faced by our project team. While the need for a long-term data warehousing effort was acknowledged in concept by GW executive management and the Board of Trustees, the project managers recognized that it is often difficult to achieve immediate widespread support for such efforts, since the perceived business value is difficult to predict and quantify. The strategy set to ensure support for the project was to secure grassroots sponsorship from a core group of management and professional level users who could directly benefit from easier access to summary-level information. This core group also provided business information requirements and tested the application. Then, while development proceeded, the project managers conducted frequent demonstrations to successively higher executive levels. Project Approach Our approach for the first phase was driven by the belief that addressing academic metrics would provide the exposure and sponsorship needed for sustaining the project. This approach has been one of the critical success factors for the project and is relevant to other higher education institutions as well. Our focus on academic metrics and our holistic approach of addressing the major milestones in a student's lifecycle at the university makes our solution unique. SDM metrics are of immediate interest to administrators and faculty alike since they directly impact GW's ability to cope with a variety of emerging challenges and are based on GW's core student data --recruitment, admissions, enrollment, and registration. Planned SDM enhancements include metrics for graduation data. Future projects include the creation of accounts receivable, human resources, and financial data marts. Technical Capital SDM data models along with best practices for architecture, database design, extract-transform-load (ETL) process design, validation, security, training, and change management constitute a significant body of technical capital for higher education data warehouse practitioners. The project team utilized several industry-standard technologies. Incremental development was based on dimensional data models specifically created for SCT's Higher Education Solution, Banner. Dimensional data models are widely accepted in the industry as the de facto standard. GW's data models are at a high level of

	abstraction and designed to naturally support high-volume queries in an efficient manner. The project team also determined early on in the project that a scalable architecture was critical for future expansion of the SDM. The team conducted a thorough analysis of tools and technologies and selected a suite of hardware and software components from market leaders with a proven track record.
Contact:	**This person has agreed to be contacted for more information on this effective practice.** Peter Barton Manager, Data Administration The George Washington University (202) 994-3413 pabarton@gwu.edu