

**A Review of the Campus Data Warehouse – Its Future Direction and Priorities
Prepared by the Institutional Research Work Group-Expanded
March 2006**

Background

At the request of the Institutional Research (IR) Policy Committee, the Institutional Research (IR) Work Group has held a series of meetings over the last several months to discuss and evaluate the Campus Data Warehouse (CDW) for the purpose of making recommendations on the future direction and priorities. This paper describes the primary observations made by the group during its deliberations and offers several recommendations on how to improve Warehouse operations and other campus efforts to more effectively develop, coordinate, and provide greater access to campus institutional data.

Group Membership and Process

The IR Work Group is a standing committee that supports the work of the IR Policy Committee. The Work Group is composed of representatives from the following units: Office of Resource Management and Planning, Student Affairs Research and Information, Information and Educational Technology, and Graduate Studies. To undertake this study, the Work Group was expanded to include representatives of several academic units (i.e. College of Letters and Science, and the College of Engineering), additional technical staff from ORMP and IET and the Office of Undergraduate Admissions. A complete list of participants is displayed in Attachment 1.

The group met four times between December 2005 and March 2006. Initial discussions focused on the history of the Warehouse and the contents of a draft paper prepared by IET Director Ray Reveles that explains the Warehouse structure, operation, and future plans (see Attachment 2). More recently, group members have also reviewed an extensive report on the Data Warehouse prepared by IET Director Debbie Lauriano (see Attachment 3). The issues raised in both of these reports have had a significant impact on the group's discussions and recommendations. This brief report does not attempt to duplicate this earlier work. Instead, it attempts to highlight the most significant factors that are impeding the development and overall effectiveness of the CDW.

Clarifying Terms

During the course of the group's discussions, it became clear that a number of the terms that are frequently used by institutional data users and the technical staff who are largely responsible for developing and maintaining data were not commonly understood. This confusion over basic terminology has is at least partly responsible for some of the confusion and ambiguity that currently surrounds the Warehouse and its future. To facilitate effective communication of basic data concepts and relationships across staff with widely varying degrees of technical expertise, the following set of basic definitions were adopted:

- **Transactions Databases** – major databases that support the daily operations of the campus (e.g., DaFIS, Banner, PPS). Data for these databases is entered by units with primary responsibility for the data. Members of these units need the ability to enter and retrieve data from the transactions database.
- **Retrieval Databases** – databases that are made up of regularly scheduled snapshots of data in transactions databases as well as specified snapshot aggregations. These databases are used by staff who need point-in-time and/or summarized data from a particular transactions database. In order to be accessible in all of the forms needed to meet user needs, the data from transactions databases must be compiled consistently in a certain way. Many transactions databases currently have retrieval databases that are well known and are quite effective (e.g., DaFIS, PPS). Other major transactions databases do not have retrieval databases (e.g., Banner) and as a result these data sets are not available in a useable form to many institutional data users.
- **Retrieval Access Tools** – Data is retrieved from retrieval databases using a retrieval tool. Because the nature of the data needed and the technical expertise of campus users varies widely, several tools are needed including: pre-defined reports (e.g., DaFIS decision support) for the most commonly needed data and one or more of the data retrieval applications (e.g., Excel, Brio, Access, SQL). Currently some retrieval databases are configured to accept the full range of retrieval tools (e.g., SIS DS) while others are not.

Note: A Transactions database and its corresponding retrieval database(s) do not necessarily have to reside in the same physical location. Retrieval databases can be maintained in one location (e.g., a campus data warehouse), housed in a decentralized way with their respective transactions databases, or a combination of both. The importance of these concepts is that these two types of databases are distinctly different and essential to the integrity, accessibility, and effective use of institutional data. Further, the location of retrieval databases and access procedures must be readily known to all potential institutional data users.

Observations

During the group's discussions, a number of fundamental observations or findings were made that eventually provided much of the support for the recommendations included in this report. Among the most significant findings were the following:

Campus Appetite for New Institutional Data Sets Continues to Grow – The campus demand for new and/or expanded transactions and retrieval databases will continue with or without the CDW. For example, MyInfoVault, Grad SMAART, and Electronic Research Administration - InfoEd are all new or planned transactions databases that address major data needs. This type of development is very likely to continue in the future. Some of these databases include plans for retrieval databases and retrieval tools for users outside the immediate units sponsoring their development. Others do not. These differences in approach highlight the need for a campus strategy to realize

the full value of such data for campus decision-makers outside the units that are primarily responsible for developing these data sets.

CDW History – As originally conceived the CDW was modeled after data warehouses at major research universities around the country. Started as a pilot program in 1998, it was intended to be a one-stop-shopping center for all types of institutional data. A major component of this initial effort was the use of a relatively sophisticated retrieval tool known as Hyperion Analyzer that allowed users great flexibility in building queries to meet specific needs. The CDW was originally housed in the ORMP and was transferred to IET on July 1, 2002

Problems with the CDW and the Coordination of Campus Institutional Data

Generally – The Work Group noted that many institutional data concerns were not necessarily directly related to the CDW. Rather, the absence of a clear plan for the CDW made it a lightning rod for a number of unresolved issues that reflect the lack of an overall plan for organizing and improving access to institutional data on campus. The following are some of the most significant issues identified by the Work Group for both the CDW and the broader institutional data issues that the CDW has come to represent:

1. The CDW did not systematically and publicly transition from being a pilot project to a permanent component of an overall campus plan to coordinate institutional data, and as a result its future and reliability are questioned by many potential users.
2. The goal of providing one-stop-shopping for institutional data was never realized even though the CDW currently contains a number of retrieval databases.
3. Hyperion Analyzer, the primary CDW retrieval tool, proved to be expensive to maintain and difficult to use, and, therefore has recently been discontinued.
4. The CDW currently faces two fundamentally different demands for institutional data: integrated retrieval databases {e.g., the Instructional Activity Information System (IAIS), and Grad System for Management of Admissions Alumni & Records Tracking (Grad SMAART)} and student data from Banner.
5. Potential users are confused by the lack of a clear understanding of the data they can find in the CDW and where they need to go to obtain institutional data that is not stored there.
6. The absence of a functional retrieval database for Banner data needs to be addressed directly and not as an add-on to a new charge for the CDW.
7. The lack of a campus entity with basic responsibility for helping IET and the CDW prioritize the development of new retrieval databases where there is proven need and guidance on the provision of adequate retrieval tools has been a major impediment to improving the overall coordination and development of high quality institutional data on campus.
8. The lack of a clear expression of the importance of building retrieval databases and their access tools into the implementation plans for new transactions databases when it is appropriate has made it increasingly difficult to realize the value of institutional data on campus.

9. Identifying and building the specific components of CDW retrieval databases in the future needs to be guided by an understanding that data users are primarily responsible for identifying useful data to collect and IET is responsible for the building the databases that deliver data.
10. The bridge between the technical aspects of databases and the policy issues and questions that data users want to explore needs to be strengthened within IET with the identification of staff who can facilitate and regularly publicize access to CDW for non-technical users.

Recommendations

The following recommendation flow from the observations noted above. They are organized by area of concern and are not rank ordered.

Campus Data Warehouse

1. **Consider Changing the Name** – The name, Campus Data Warehouse, may carry too much historical baggage. Once its primary responsibilities are established, a new name might be appropriate and beneficial. For purposes of this report, the name will remain unchanged.
2. **CDW Staffing** – Assigned staff within IET should be full time like other major databases in IET and include an analyst/ombudsmen.
3. **Primary Responsibility of CDW Staff** – Build new integrated retrieval databases (e.g., IAIS) where the need has been identified, and maintain other existing retrieval databases that provide certain types of access not available elsewhere (e.g., selected data for employees and students currently in the CDW)
4. **First Priority of CDW Staff** – Ensure the continuing availability and integrity of existing retrieval databases in the CDW.
5. **Entity Responsible for Oversight of future CDW development** – Institutional Research Policy Committee, with the assistance of the IR Work Group - Expanded.
6. **Advisory/User Group to CDW**– Expanded Institutional Research Work Group
7. **Near Term Priorities** – (1) Develop a new integrated database focused on graduate students with emphasis on financial support; and (2) identify/develop, and provide support for a specified number of retrieval tools, beyond pre-defined reports, to take the place of Hyperion Analyzer.

Banner Data

8. **Missing Components** – The Banner transactions database needs to be supplemented with a comprehensive retrieval database and retrieval tools
9. **Internal IET responsibility** – IET has a set of positions dedicated Banner and a set dedicated to CDW. Responsibility for building the Banner retrieval database and installing appropriate retrieval tools should not be assigned to IET CDW staff.

User Support

10. **Ombudsman** - Every major retrieval database, or group of such databases, should have an analyst/technician ombudsmen to help campus users obtain the data they need
11. **Retrieval Tool Support.** - The campus should investigate providing user support for the most commonly used retrieval applications (e.g., Brio).
12. **Marketing Plan** - IET should develop and regularly disseminate a clear description of major data retrieval databases, and retrieval tools that are available

Campus Institutional Database Policies for the Future

13. **Retrieval Databases and Tools** - Every major transactions database should have a retrieval database and data retrieval tools that reflect the needs of most users
14. **Retrieval Database Construction** – Content should represent best practices and standards of measurement defined by primary users (functional responsibility). The relationship between primary users/ owners and IET staff should be characterized by the primary users providing the best data to the technical staff responsible for building the retrieval database.
15. **Budgeting for All Three Components** - When a new transactions database is developed, the project budget should include the creation of a retrieval database and appropriate retrieval tools or a plan for providing these components in the future

Attachment 1 – Membership - IR Work Group – Expanded

Attachment 2 – Draft report on the Campus Data Warehouse prepared by IET Director Ray Reveles, December 2005

Attachment 3 – Student Information System and Data Warehouse Project Report prepared by IET Director Deborah Lauriano, November 20, 2002

**Membership
Institutional Work Group – Expanded**

Work Group

- Bob Agee, Resource Management and Planning
- Helen Paik, Resource Management and Planning, Institutional Planning and Analysis
- Dave Shelby, Information and Educational Technology
- Ray Reveles, Information and Educational Technology
- Elias Lopez, Student Affairs Research and Information Office
- Yuhang shi, Graduate Studies

Expanded Membership

- Eric Rothgarn, Resource Management and Planning
- Julie Saylor, Resource Management and Planning
- Brian Alexander, Undergraduate Admissions
- Dann Trask, College of Letters and Science
- Donna Davies, College of Engineering
- Mary Brown, College of Engineering