

Topic Title **Updating the Data Mart: Background**

A Basic Primer on Data Marts

This handout is intended to prepare you for training in updating the Institutional Research Data Mart (DMIR). After reading this handout, you will know what a data mart is and its context in a reporting environment.

What is a data mart?

To understand what a data mart is, you must first know a little about data warehouses. The concept of data warehousing first appeared in the late 1980s in articles published by Bill Inmon and others. Essentially, a data warehouse involves a **systematic approach to gathering, organizing, and storing data**—generally from internal production sources but also from external providers—to create a **definitive source of information for analysis and decision support**.

Information systems, such as the PeopleSoft Student Administration system used at Santa Clara University (SCU) and many other higher education institutions, are built to handle **transaction processing** such as admitting and enrolling students, administering financial aid, managing student accounts, assigning faculty to classes, tracking students' progress toward completing their programs of study, and many similar operations. They utilize databases that are necessarily complex and are designed to hold large amounts of very detailed data that are frequently updated. The PeopleSoft database has hundreds of tables with thousands of fields and millions of rows of data that are often "effective dated." Effective dating allows users to see, for example, when a student added a particular class and how much the student now owes the university.

However, it isn't easy to query these data to produce consistent reports that allow administrators to see, for example, the number of enrolled students aggregated by gender and ethnicity. To create these kinds of **managerial and strategic reports** requires a data warehouse, which is a database with simpler data structures optimized for reporting and analysis rather than transaction processing.

A data mart is like a data warehouse but with a **limited scope**. A data warehouse usually contains enterprise-wide data, whereas a data mart stores data for a particular **business unit or function**, or it is designed to answer **specific questions**. For example, an admissions data mart

Topic Title **Updating the Data Mart: Background**

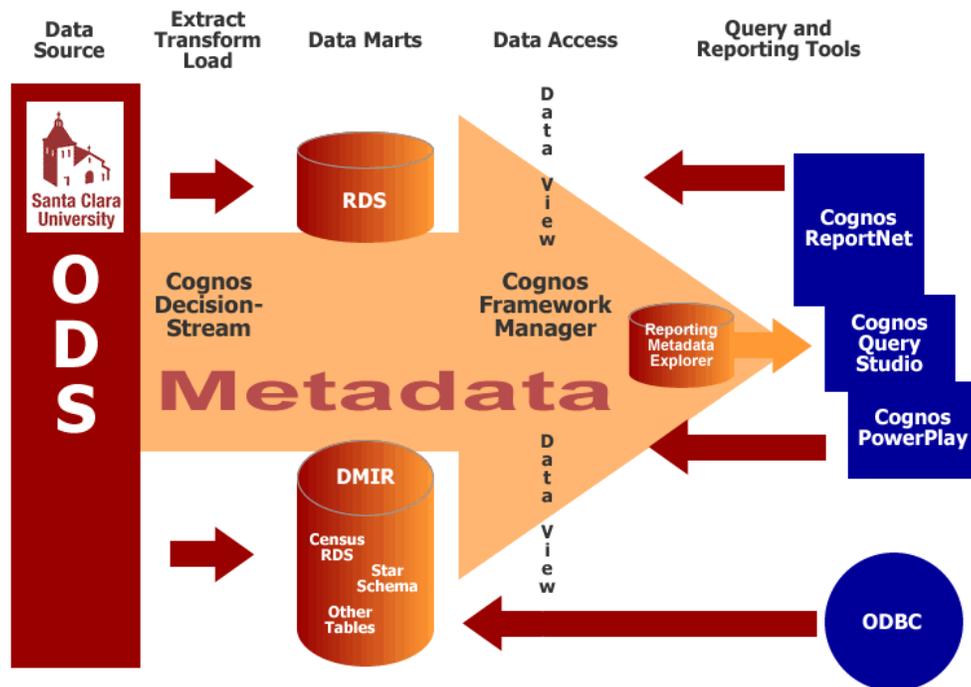
would have data about prospective students, the subset of those prospects who are recruited, and the further subset who actually apply to the university. These data would include descriptive information like biographic and demographic data (sometimes referred to as “dimensions”) as well as quantitative information like SAT scores and high school GPAs. It wouldn’t have data related to faculty or courses offered. Data marts are usually populated using a **top-down approach** in which data flow from source systems into an enterprise data warehouse and from there into data marts.

SCU doesn’t yet have a data warehouse, so DMIR is populated with data extracted directly from the PeopleSoft Student Administration system. Because Institutional Research analyzes and reports data for many business units that don’t yet have data marts, the scope of DMIR is broader than the typical data mart.

Topic Title **Updating the Data Mart: Background**

The SCU reporting environment

The diagram below illustrates the way data are moved from the source database (a copy of the PeopleSoft transaction database referred to as “ODS” or “operational data store”) into data marts. The process of extracting data from the source and inserting it into the target is referred to as “ETL,” which stands for “Extract, Transform, and Load.” We use DecisionStream from Cognos to accomplish this.



In addition to DMIR, we have a data mart named “RDS” (“reporting data store”). While both DMIR and RDS are data marts with similar structures (i.e. they have many of the same tables and fields), they serve different purposes. DMIR holds point-in-time data such as census enrollment data for many years; RDS data are refreshed nightly. If a user wants a report based on data from the Fall 2004 census date, she would query DMIR. If she wants a report based on current data, she would query RDS.

The data in the data marts are presented to users through a data view that is managed using another Cognos product, Framework Manager. The database administrator uses this tool to join together various tables in a data mart and to give user-friendly names to tables and fields.

Topic Title **Updating the Data Mart: Background**

For example, the ADM_APPL_NBR field in the ADM_APPL_DATA_DIM table might be presented to the user as the “Application Number” in the “Admissions Data” table.

Users have a few tools from Cognos for reporting and analyzing; the choice of tool depends on the type of activity in which the user wants to engage. If she simply wants a quick query of the database to get a count of the number of student applications we’ve received, she would use Query Studio. If she wanted a more complex report suitable for distribution, she would use ReportNet. And if she wanted to analyze highly aggregated data and have the ability to “drill down” into the data to explore a particular area in greater detail, she would use PowerPlay. In Institutional Research we also have ODBC (“open database connection,” a standards-based method of access databases) access to DMIR, which means we can bypass Framework Manager’s data access layer using any tool that is ODBC-compliant, including SPSS and Microsoft Access and Excel.

The reporting environment illustrated by this diagram may be summarized as follows: Data from a data source are extracted and loaded into data marts and presented to users who access them using a variety of tools.