

The Massachusetts Data Warehousing Initiative

Jen Goree <http://www.doe.mass.edu/infoservices/dw/>



What is a data warehouse?

A data warehouse is a repository of data from many sources. Unlike a typical "transactional" system used for day-to-day operations, a data warehouse is structured to maintain large amounts of related, historical data for analysis and reporting. A data warehouse provides for easy reconstruction of "snapshots" of historical data, as well as the ability to link such snapshots over time using certain criteria.

Why do we need a data warehouse?

Improving educational performance and accountability depends on understanding the relationships among areas such as curricula, assessments, special programs, teacher qualifications, program spending, discipline incidents and attendance. With our current systems it is difficult to cross-reference student-to-teacher, student-to-course, or student-to-program information. Without the ability to link data, Department and district decision-makers cannot leverage the full potential of the information we collect and process. The data warehouse is a "longitudinal data system" capable of linking student, teacher and financial information over multiple years, across multiple schools and districts. Using the warehouse, Department and district decision-makers can take key metrics from multiple areas and analyze them in a single view. For example:

- What are the results of program participation in terms of achievement or graduation rates?
- Which programs targeted at which student population have the greatest impact?

Who is developing the data warehouse?

The Department is working in partnership with Cognos Corporation, an organization that specializes in large-scale data management systems. Cognos was selected through a competitive procurement process. Similar Cognos educational warehouse systems are operational today in 12 states across the country. Local districts have the potential to play a major role in populating the warehouse with historical data from their administrative systems. The accuracy and consistency of the data in a district's administrative systems will largely determine the accuracy and consistency of the district's warehoused data.

Will this system replace a district's or a school's current student information system (SIS)?

No. The data warehouse is capable of drawing data from many sources, including local SIS applications. It is designed to easily load archive-ready data "in bulk" at regular intervals and to provide enhanced analysis and reporting on interrelated data from multiple sources and time periods. The data warehouse is not meant to manage current day-to-day transactions.

Where will the data in the warehouse come from?

The Department has loaded MCAS/MEPA and SIMS data from 2002 to 2006 into the warehouse. A number of pilot districts have loaded local data, including student, course, staff, grades, GPA, and local assessments. Additional districts will have the capability to load data from school/district administrative systems over the next several months.

In later phases, more districts will be able to load additional student, course and teacher data. The warehouse may also be expanded to include discipline and attendance detail, financial, and program data. The primary source of warehoused data will be school/district administrative systems, including SIS, Gradebook, HR and Finance systems. Districts can periodically upload data to the warehouse as needed. The data warehouse will provide districts with the ability to maintain large amounts of integrated, historical data for analysis and reporting.

Who will have access to my district's warehoused data?

All users can access SIMS and MCAS data in aggregated form, as with public profile reports. District Directory Administrators will also be given the capability to authorize district users to view more confidential data.

Only those Department staff that have a need for the data to fulfill their professional responsibilities will be given access to confidential student SIMS and MCAS data, in line with current Department policies. The Department is also working to develop a policy for district data access that addresses the needs and concerns of superintendents and Department officers. All data warehouse users will be required to sign the Data Warehouse Data Access Policy.

For additional information on access to confidential student information see Reporting Guidelines Frequently Asked Questions at <http://www.doe.mass.edu/infoservices/data/sims/simsfaq.asp>.

What tools will districts use to analyze and report on the historical data in the warehouse?

Authorized users may access the warehouse analysis and reporting tools online, through the Security Portal. The primary reporting tools are Cognos Connection, the main portal and report viewer; Cognos PowerPlay, a cube-based analysis and reporting tool; and Cognos Report Studio, a report-authoring tool. The Department will provide pre-defined reports and analysis cubes, and district staff will have the ability to define their own reports or request additional cubes.

What training will my staff need, and what training resources are available?

The Department has currently identified three general levels of training:

End Users: Users who wish to access predefined reports and analysis cubes are expected to need minimal training. The Department has licensed self-paced, computer-based training from Cognos that may be adequate for most end users. Cognos has also worked with the Department to develop a half-day training course that may also be offered to potential end users.

Report Authors: District and Department users who wish to author their own reports will be required to attend ReportStudio training. Cognos offers a 2-day ReportStudio training for general audiences on a regular basis. Additionally, the Department and Cognos have developed a one-day course for report authors that may also address this requirement. The Department will not grant access to ReportStudio to users who have not completed training.

Data Loading: For districts that opt to load local data, district technical staff responsible for producing data extracts will need to become familiar with the extract formats and procedures for generating and uploading local data extracts. The Department will provide support and training as needed on a district-by-district basis, and will work with district technical staff and SIS vendors to produce and load the necessary extracts. However, districts should expect to devote considerable time and resources to producing the initial loads. For the current local-data-loading pilot districts, time from extract start to first successful load ranged from three to six months.

Will this system replace TestWiz?

The data warehouse and reporting system is expected to greatly expand on TestWiz's analysis capabilities. Currently there are no plans to phase out TestWiz until the Department is confident that it can provide access and support for the data warehousing system.

My district is considering incorporating a new assessment into our instructional plan. Should I wait to see if it is supported by the data warehousing system before I decide to use it?

In general, no. The data warehousing system is a tool intended to support policymaking and instruction. When deciding whether to adopt a new assessment, the primary concern should be whether or not it will improve educational outcomes for students, not whether or not it will be compatible with the tool. With that in mind, we do recommend that schools and districts consider assessments with the following characteristics:

- a. Can produce results in electronic format. Assessment data must generally be exportable into flat files in order to be loaded. Paper-only results may create cumbersome manual-entry tasks for district staff.
- b. Can incorporate the use of state identifiers (SASID, district/school codes) into result files. This will greatly simplify the task of linking assessment results to existing data.

What is the scope and timeline of the pilot project?

The pilot project is currently in its first phase. The goals for this phase are:

- Deploy reporting tools on SIMS and MCAS data (2002-2006) to approximately 35 districts.
- Enable seven districts to load local data, including student, course, staff, GPA, and local assessments.

The second phase, expected to begin during the 2006-07 school year, will enable the remaining pilot districts (beyond the 'Early Adopters' listed below) to load their own local data.

How were the current pilot districts selected?

The current pilot districts were awarded competitive grants based on their level of preparedness and need. One grant project, serving 7 districts, is in the second (and final) year of its grant project and will be loading local student, assessment, course, staff, grade, and location data. The six remaining grant projects, serving approximately 30 districts, will have access to reporting tools on SIMS and MCAS data, and will begin the process of loading local data in the first half of this year.

Can my district sign up to be a pilot district?

While the districts for the initial pilot phase are fixed, the Department will be discussing how to expand the project beyond the current pilot districts at the conclusion of this phase (summer 2007). The methods and procedures for expanding the pilot will depend on the results of the current phase, as well as the amount of funding allocated to this project by the Legislature. We will post more information as it becomes available.

Should my district consider acquiring its own data warehousing system?

While we cannot prevent districts from doing so, the Department strongly recommends that districts consider the long-term benefits of a statewide data warehousing system. Most district-level systems require an annual per-student fee that can result in heavy long-term costs.

The statewide system, on the other hand, expects to provide a base level of functionality at minimal cost to districts. Extended functionality, including loading of local data, involves upfront costs devoted to building district capacity and training, and complete system ownership after an initial development period — in other words, the long-term costs are expected to be far lower due to district self-sufficiency.

In addition, the Department expects districts to benefit greatly from using a standardized system. Specific reports, professional development, and best practices can be shared easily among districts using a common platform.

Who are the districts involved in the pilot?

The pilot districts are listed below, by grant project (N.B.: Some entities listed below comprise multiple districts):

CRISTAL Partnership (Early Adopters):

Northampton (lead agency)

Amherst-Pelham

Franklin County Technical School

Frontier

Gill-Montague

Mohawk Trail

Pioneer Valley

ACCEPT Collaborative:

J. P. Keefe Technical School

South Shore Charter School

Ashland

Holliston

Hopkinton

Natick

Dover-Sherborn

Franklin

Medfield

Medway

Millis

Sudbury

Community Day Charter School

DEEP Project:

Wareham (lead agency)
Barnstable Horace Mann Charter School
Dartmouth
Bourne

"OSN" Grant Project:

West Springfield (lead agency)
Easthampton
Gateway

CRISTAL LITE Partnership:

Greenfield (lead agency)
Ralph C. Mahar Regional School
Orange
Ware
Union 28

Worcester Public Schools

Holyoke Public Schools

Winchendon Public Schools

Where can I get more information on the data warehouse?

For additional information on the data warehouse project, check for updates on the public web page at:
<http://www.doe.mass.edu/infoservices/dw/>.

Pilot participants are invited to join the Data Warehouse workgroup on MassONE. Project documents are available through the workgroup. To join the workgroup, register for a MassONE account (<http://massone.mass.edu/administrators/register.html>) and then email your request to the data warehouse project team at datawarehouse@doe.mass.edu.