How Can My State Benefit from an Educational Data Warehouse?

September 2007

By Terry Bergner and Nancy J. Smith, Ph.D., Data Quality Campaign

Imagine if, at the start of the school year, a teacher could have detailed information about the academic history of every student in her or his classroom. This is possible if the teacher can log on to a Web site that provides access to an educational data warehouse. The teacher would see not only several years of state assessment results, but also enrollment history, demographics, program participation, discipline records, and schedule and transcript data. With this reliable and easy-to-access background information, the teacher could design an appropriate academic approach to meet the needs of each individual student and the class as a whole.

In addition, when districts have agreements about accessing a shared data warehouse, school administrators and teachers can use the information stored in the warehouse to properly place a student who moves to a new district in the appropriate classes and/or programs efficiently and quickly. These are just a few examples of how a well-designed student data warehouse can be a key factor in improving student achievement and teacher quality.

According to the 2006 Data Quality Campaign (DQC) survey,[1] 26 states have designed and built or upgraded their data warehouses or are in the process of doing so. State policymakers and educators need a data system that not only links student records over time and across databases, but also makes it easy for users to query those databases and use up-to-date reports to adapt to the unique needs of their students.

By using unique identifiers and linking the records from several areas (e.g., student and staff records), states have made new resources available to teachers and administrators, and the possibilities for improving

> ### Highlights
>
> In this brief, find out more about:
>
> ▶ What an educational data warehouse is;
>
> ▶ The benefits of developing and using a data warehouse; and
>
> ▶ Design and implementation recommendations from leading states.

educational outcomes are overwhelming. The engine behind this is the educational data warehouse. Robust data warehouses are vital for providing useable data to policymakers and educators, and many states and school districts are investing in them.

Many larger districts are ahead of their state education agencies (SEAs) in the development and use of educational data warehouses. SEAs just starting down the road can learn from those entities and develop methods for sharing data between the respective warehouses in an efficient, interoperable manner. However, the vast majority of small- to medium-sized districts will benefit from such technology only if SEAs build, manage and allow access to the data warehouse, ensuring economies of scale for all stakeholders. Therefore, SEAs must plan and budget for data warehouses that are accessible to policymakers, parents and educators statewide as a natural extension of the robust longitudinal data systems outlined in the DQC's 10 essential elements. (See box, page 3.)

---

[1]Data Quality Campaign/National Center for Educational Accountability, *2006 Survey of State Data Collection Issues Related to Longitudinal Analysis.*

The educational data warehouse can generate or feed a comprehensive system of both standardized and customized reports and analytical tools that answer questions faced by policymakers, district administrators, local educators and parents. Implementing and using a data warehouse, along with reporting and analysis tools, can enable educators and policymakers to effectively use the vast amounts of data collected in the state's data system.

Discussions about data warehouses and reporting and analysis tools so often occur at the same time that it is easy to think that they are one and the same. Although these two components of a robust data system go hand-in-hand, they require separate types of technology and staff expertise. A data warehouse is, in essence, a storage facility for many datasets culled from a variety of source files, such as student enrollment, program participation, graduation, state-level test data, teacher data and financial data. Reporting and analysis tools, however, are essentially the software programs written to calculate the statistics that stakeholders need to evaluate the performance of a student, school, district or state and produce reports (electronic or print) that answer stakeholder questions. A separate implementation brief about reporting and analysis tools will be published alongside this brief about data warehouses.

The DQC interviewed staff from three states (Delaware, Maryland and Wyoming) to ascertain how each state approached the design and development of its educational data warehouse and to gather recommendations to share with other states. These three states were selected because of the varying ages of their data warehouses and their history of data collection and use.

What Is an Educational Data Warehouse?

Though the term "data warehouse" may mean different things to different people, for the purposes of this brief, an educational data warehouse is a storage facility, built and maintained by an SEA, where detailed and reliable educational data from several areas that affect student achievement are stored and integrated. These data then can be used to produce a variety of reports for a wide range of audiences, from the general public to the individual teacher. Because several years of data are represented and data have been joined together from many databases, these data can be analyzed and used in new ways.

When data reside in separate "silos," it is often impossible to conduct new levels of detailed student- or teacher-specific analyses to inform policymaker and educator decisions. Without implementing an interoperable solution to feed a comprehensive data warehouse, the student, staff or even financial records often cannot be combined easily with each other or with data from other areas to provide a comprehensive picture of achievement or be examined longitudinally to measure progress.

Benefits of Developing and Using a State Data Warehouse

The business of education has changed. A few years ago, a mere handful of SEAs had individual student records and an operational data warehouse. Although SEAs may have understood the value of and need for a data warehouse for both reporting and improvement purposes, it was something they could only hope for in the future. Because of the quest for outcomes-based accountability spurred in part by new federal and state mandates for complex, longitudinal data analysis (such as the data reports demanded by the No Child Left Behind Act), yesterday's want is today's necessity.

In addition to facilitating more efficient reporting, state data warehouses are able to provide access to

▶ Because of recent improvements to its data warehouse that allow it to link student, assessment, staff and facilities data, **Wyoming**'s Department of Education is finding the analysis needed to provide complex longitudinal reports required by No Child Left Behind a fairly manageable, albeit time-consuming task.

▶ A central tenet of the **Delaware** data warehouse is to use data for informed decisionmaking at all levels, from the administrator to the parent.

data and data analysis quickly and efficiently to improve resource allocation, instructional programs and techniques and to foster a culture of continuous improvement.

10 Essential Elements of State Longitudinal Data Systems

1. A unique statewide student identifier that connects student data across key databases across years
2. Student-level enrollment, demographic and program participation information
3. The ability to match individual students' test records from year to year to measure academic growth
4. Information on untested students and the reasons they were not tested
5. A teacher identifier system with the ability to match teachers to students
6. Student-level transcript information, including information on courses completed and grades earned
7. Student-level college readiness test scores
8. Student-level graduation and dropout data
9. The ability to match student records between the P–12 and higher education systems
10. A state data audit system assessing data quality, validity and reliability

Fundamentals in Designing State Longitudinal Data Systems

In addition to the 10 essential elements, states need to ensure that they take into account the following fundamental concepts in the construction of their longitudinal data systems:

1. **Privacy Protection** — the assignment of unique student identifiers and guarantee that personally identifiable data do not become available to data users
2. **Data Architecture** — the documentation and enforcement of rules regarding how data are coded, stored, managed and used
3. **Data Warehouse** — a repository of data, such as student, staff, curriculum, facilities and finances
4. **Interoperability** — the ability of different software systems from different vendors to share information without customized programming or data manipulation
5. **Portability** — the ability to exchange student transcript information electronically across districts and states or between P–12 and postsecondary systems
6. **Professional Development around Data Processes and Use** — the training of people charged with collecting, storing, analyzing and using data
7. **Researcher Access** — the capacity and willingness to share data with researchers while meeting federal and state privacy regulations to enable research and evaluation studies

Design and Implementation: Recommendations from Leading States

As states begin to think about the design and implementation of an educational data warehouse, they can benefit from recommendations from states that have gone down this path before them.

1. Know the conditions and purposes of establishing a warehouse.

The availability of SEA staff and the expertise of existing staff members will dictate how much of the implementation process is conducted in-house versus contracted, as will the availability of financial resources.

> ▶ **Delaware** chose to work with a "What Why How and How Often" format to build its infrastructure. The state established ground rules and then put it all together with a common data dictionary.
>
> ▶ **Maryland,** in partnership with the Center for Technology in Education at John Hopkins University, has developed a standardized Web-based Individual Education Program (IEP) for special education students. Data entered into the IEP data collection system online are fed directly into the special needs student record at the state department of education. These data are submitted daily, and the numbers and IEP details for their students are available for district educators to see and use at all times. Maryland is building this special education system as a prototype for an "all-student" data warehouse, which is currently under way.

Tailor the warehouse to the specific state's legislative mandates, culture and expectations of education data. Know what the purview of the work should cover realistically.

2. Identify the scope of the project, have realistic expectations, build a strong project plan and stick to it.

Ensure that the budget meets the needs of the plan, revisit the plan often and avoid reinventing the wheel. Much time, effort and money can be saved by learning from states that have implemented a similar project

successfully. Involving key stakeholders, both within the SEA and outside, at the beginning of the design process will assist in developing a reasonable scope from the beginning that all parties agree to. Early buy-in from outside stakeholders is essential for a well-managed project.

> ▶ **Delaware** suggested picking a purpose and a starting point and later branching out from there, yet staying true to the original purpose of the project and completing the initial tasks before expanding.
>
> ▶ **Wyoming** stressed the need to have a management-level expert and champion at the SEA and to build capacity in-house.

3. Generate realistic estimates of time and cost.

Recognize that time and costs are dependent on the variables and scope of the warehouse. These are some of the questions that need to be answered before a realistic answer can be given:

▶ What is the scope of the project?

▶ Is there an external infrastructure in place to collect data electronically?

▶ At what point are you starting? Do you currently have individual student and staff records?

▶ How much hardware and software will need to be purchased?

▶ Are the current data comparable?

▶ Are there adequate staff resources at the SEA? How much of the work will have to be contracted out?

In most cases, the simple answer to the time and money question is years and millions. However, the price also needs to be viewed in terms of the cost savings that the new technology will bring. Through the process of documenting existing data collections, storage and reporting processes, states often identify ways

to make their processes more efficient (such as reducing the number of data collections, requiring fewer staff to oversee specific data collections or reporting activities, and integrating and automating the merging of files from different program areas using standards and common definitions).

> ▶ **Delaware** contracts out $1 million per year for database administration, programming and Web services. In Delaware, SEA staff numbers are limited by legislation. They also have a four-year hardware life-cycle plan for which they set aside $250,000 each year.
>
> ▶ Noting that it had a small number of students compared to other states, **Wyoming**'s basic investment in the infrastructure at the SEA was $150,000 in hardware, $180,000 in software and licenses, and $40,000 for annual support. However, the original data warehouse implementation plan did not allow for true longitudinal data tracking, and the state has since taken on a project to expand in this direction.
>
> ▶ **Delaware** began work on its educational data warehouse in 1984 with longitudinal assessment records; Maryland began in 2001 with its special education system; and Wyoming began its original data warehouse implementation in 1989, with its addition in 2004. All still are growing and have plans to expand their systems in the future. This year alone, Delaware is adding transcripts, schedules, Advanced Placement and SAT scores to its data warehouse and is planning to build a link between its P–12 and higher education systems.

A state that currently has only aggregate student data at the school-building level and is just beginning to collect individual student records with a unique student identifier will need much more time than one with a student information system already in place.

4. Tap into federal, state and local funding.

Local, state and federal funds are all possible sources for building and maintaining a data warehouse. All three states began mainly with funds from the state and over time applied for and received additional funding from federal grants. The annual monies needed for maintenance usually have come from the state.

5. Include stakeholders/users of the data in the planning process.

Stakeholders from every level in the education system, including staff from within the SEA, district and school representatives, parents, legislators, policymakers, researchers, and vendors need to be involved in developing and constantly shaping the data warehouse. The importance of including stakeholders, along with developing a method to collect and evaluate their needs and ideas, from the beginning of the project and after the data warehouse is deployed cannot be overstated. Stakeholder involvement in the design phase is needed to educate all parties about what a data warehouse can do, facilitate widespread buy-in for the tool, and create a common long-term vision of data collection and use throughout the state. However, it also is important to manage time and participation so that the team identifies solutions and action items on a regular basis rather than discussing situations endlessly.

> ▶ From the beginning, **Maryland** had a formal working stakeholder group that helped to outline the data needed for federal and state reports and develop the data collection and user access procedures. Maryland's online IEP system has a "Feedback" button, and the SEA has held monthly workgroup meetings since April 2006 to share, discuss and address statewide online IEP issues.
>
> ▶ **Delaware** gathered input from several focus groups. The state worked, and continues to work, side-by-side with its teachers unions and district and school administrator groups. Although it is informal, Delaware's evaluation procedure is ongoing and includes all stakeholders and SEA staff.

6. Document existing data collections and reports.

Conduct an analysis of all existing data collections, and eliminate all duplicate efforts to collect the same data. Information about reports required by state and federal agencies, and when are they are due, need to be documented. To facilitate efficiency and ensure quality, develop common data definitions and a corresponding data dictionary. Documenting processes and

data collection activities usually results in streamlining activities and eliminating redundant data collections, storage and reports. The data warehouse provides a vehicle for combining data from different collections in one place, thereby making the production of common reports more efficient.

7. Address security issues up front.

In today's age of advanced technology and built-in security features included in all software, system security is not an insurmountable problem; however, the security issue cannot be ignored or underestimated. Depending on the user, the type and level of information available will vary greatly. With the exception of the public area of the SEA Web site, security is handled through the use of user names and passwords that permit an appropriate level of access to each user, according to his or her need to know. All parties must be informed of the safeguards in place and have confidence in the system.

> ▶ Any parent in search of new housing in **Delaware** is able to access the public portion of the state Department of Education's Web site to see what school his or her child would attend and to view a "Report Card" for the school. At the most secure level, a parent in Delaware can have access to his or her own child's record, which includes information about schedules, grades, state assessments, attendance and discipline.
>
> ▶ **All three states** discussed the need to address and overcome the issue of confidentiality from the beginning — from the building of a unique student identifier to the linking of staff and student records.

8. Ensure adherence to federal and state privacy laws.

By law, federal guidelines must be followed and adhered to when addressing confidentiality in reporting. State-specific guidelines may be stricter but can never fall below the federal Family Educational Right and Privacy Act (FERPA) guidelines.

With knowledge of the safeguards in place and the building of trust among all stakeholders over time, the

barriers this issue seems to bring can be overcome (see DQC FERPA guide on building/using state longitudinal data systems while protecting privacy).[2]

9. Create and implement training and professional development.

Rules, procedures, data collection calendars, policies, maintenance and training plans have to be developed, agreed upon and put in place as the design of the SEA data warehouse proceeds.

Training is a mandatory component of every project plan. The extent and depth of training will vary depending on the audience. The public portion of the SEA Web site that disseminates the information contained in the data warehouse should be simple in design and self-explanatory. Training in the collection and use of data can occur in many forms and needs to be tailored to the specific audience (e.g., district administrators, school principals and SEA staff). A primary factor in measuring the success of a data warehouse always will be how well its users interpret and use the data. States typically develop manuals, hold classes for the various types of users, offer training and support on the Web, and operate a phone or e-mail help line. Most states use a combination of methods, but regardless of the method of delivery, training will be ongoing because of modifications to the data warehouse system and staff turnover.

10. Create and conduct evaluations regularly.

Regardless of the stage or age of the educational data warehouse, evaluation activities must be ongoing. A formal evaluation plan to ensure that the current and future needs of SEA and local education agency (LEA) staff — as well as those of all other stakeholders — are being met is necessary and can help to guarantee the continued success of the data warehouse. An

---

[2]Data Quality Campaign, *Maximizing the Power of Education Data while Ensuring Compliance with Federal Student Privacy Laws: A Guide for Policymakers,* 2007.

▶ In **Delaware,** all training is hands-on and is provided as needed to SEA and LEA users.

▶ In **Maryland,** all Web-based data collection systems have online training that works well. For its special education data system, the state has two training sites and always offers phone support.

▶ **Wyoming** provides data collection trainings via video conferencing over the state's Wyoming Equality Video Network (WEN Video), from which videos are created for later viewing.

evaluation plan should be a part of the original project management plan and should include a way to measure the degree of success of every facet of the data warehouse.

11. Ensure funding, and schedule for continual maintenance.

Investing in data systems is not just a one-time cost. Data systems require continual upgrades in both technology and the variables to be collected. As states and the U.S. Department of Education change their reporting requirements and accountability systems, the data infrastructure will need to change as well.

These changes will require funds for staff, equipment, training and other needs identified through ongoing evaluation processes. The burden of funding the maintenance of an SEA data warehouse most often will fall on the shoulders of the state.

12. Assess the role of vendors, and create a request for proposal (RFP) to gauge vendor interest.

Vendors are critical partners in the development of state longitudinal data systems. Based on their experiences across many states, districts and, possibly, other industries, they may bring valuable and innovative solutions to the table.

Although many states try to build as much in-house capacity as possible, most states need to contract out some of the data warehouse development process to vendors. Before the vendors are invited to the table, though, the SEA and LEAs should work together to thoroughly define the scope of work based on the use of standards and common definitions and define the specific role of the vendor. A well-written and properly vetted RFP should allow the vendor to become a partner in the resolution instead of another problem to be addressed.

Conclusion

Although the planning for, and the construction and maintenance of, an educational data warehouse requires a long-term commitment by the state agency and stakeholders, as well as long-term funding, the resulting improvements in educational outcomes are well worth the investments. Instead of needing a large cadre of data analysts and programmers to respond to frequently asked questions and produce a plethora of customized reports, a data warehouse allows the SEA to save resources and time by having a smaller cadre of analysts and programmers, combining large

amounts of data, and automating reporting processes. By providing role-based access to the data and standardized reports and analyses, the data warehouse allows district administrators and principals to use time that previously had been spent building their own spreadsheets and creating reports for their own schools and teachers to develop interventions and enhance programs. In short, the data warehouse provides a way to streamline data storage, analysis and reporting activities at the state and local levels.

State contacts:

Delaware Department of Education

Robin Taylor, Associate Secretary, Assessment and Accountability, rtaylor@doe.k12.de.us

Maryland Department of Education

Carol Ann Baglin, Assistant State Superintendent, cbaglin@msde.state.md.us

Ned Featherston, Analyst, nfeatherston@msde.state.md.us

Robin Gross, Consultant and Project Manager, robin@tfrservices.com

Wyoming Department of Education

Teri Wigert, Director, Technology, Careers, and Data, twiger@educ.state.wy.us

Meredith Bickell, Technical Services Supervisor, mbicke@educ.state.wy.us

Vince Meyer, Data Services Supervisor, vmeyer@educ.state.wy.us

Dave Wilheim, Database Administrator, dwilhe@educ.state.wy.us

Data warehouse vendors used by states as reported on 2006 Data Quality Campaign survey:

Bearingpoint/KPMG

Cognos

Computer Task Group

Deloitte & Touche

eScholar

Fujitsu, Inc.

Infinite Campus

Measured Progress

Pearson

Public Consulting Group

Tetra Data

Triand

Veridian

In addition, some states have built the data warehouse in-house.[3]

————

[3] The Data Quality Campaign does not endorse any vendor; this is a list of vendors used by states as displayed on the DQC Web site (www.DataQualityCampaign.org/survey_results/vendors_2006.cfm).

DATAQUALITY
CAMPAIGN
Using Data To Improve Student Achievement

www.DataQualityCampaign.org

The Data Quality Campaign is a national, collaborative effort to encourage and support state policymakers to improve the collection, availability and use of high-quality education data and to implement state longitudinal data systems to improve student achievement. The campaign aims to provide tools and resources that will assist state development of quality longitudinal data systems, while providing a national forum for reducing duplication of effort and promoting greater coordination and consensus among the organizations focusing on improving data quality, access and use.

The Data Quality Campaign has 14 managing partners and numerous endorsing partners. For the list of partners and more information, please visit www.DataQualityCampaign.org.

The CCSSO is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress and the public.