TEXAS STUDENT DATA SYSTEM

Lessons Learned: Phase 2 Prototype Development and Technical Notes

Zeynep Young
Sharon Reddehase
Ed Comer

January 31, 2011

This document and the project described herein was led by the Texas Education Agency and supported by the Michael & Susan Dell Foundation. The project was conducted by Double Line Partners, LLC.

Table of Contents

Background

The Texas Student Data System (TSDS)[1] is a multiyear initiative led by the Texas Education Agency (TEA) to (1) build a platform to deliver relevant and actionable data back to educators to improve student performance; (2) alleviate the data collection and submission burden on school districts and improve data quality; and (3) integrate key data into TEA's P–20 data warehouse to better understand students' college and workplace readiness.

In 2009, the Michael & Susan Dell Foundation (MSDF) sponsored an effort by Double Line Partners to design, develop, and document a prototype for a set of educator dashboards and a data store to drive the educator dashboards while also supporting TEA's data collection requirements. The goal was to provide timely, accurate, and actionable information in the form of easily understandable educator dashboards to improve student outcomes.

Though not without its challenges, Phase 1 of the prototype development[2] proved the feasibility of the key components of the solution: [3]

- Derivation of a reference model for relevant district data, known as the Canonical Data Model (CDM). The CDM forms the basis for an XML-based data interchange standard and for the logical data model of the data store, called the District Connections Database (DCD).
- Application of the CDM data standard to input raw data from a single district, pulling from the district's Student Information System (SIS) and other electronic sources.
- Definition of a set of high school metrics using data already available at the district level and based upon predictive indicators of student performance identified in academic and best practices research.
- Development of a prototype of high school student and campus metric dashboards that are populated from the DCD.
- Demonstration that a subset of the yearly PEIMS (Public Education Information Management System) data collection required by TEA could be generated from the DCD.

This document outlines experiences and lessons learned during Phase 2 of prototype development. In the spirit of sharing and collaboration, the document focuses on key conclusions and issues relevant to others in the education community involved in or contemplating the development of student performance dashboards and their underlying data systems.

---

[1] Information regarding the TSDS project may be found at www.TexasStudentDataSystem.org.

[2] The Phase 1 dashboard and data standard documentation may be found at www.districtconnections.com/Phase1.

[3] *Lessons Learned: Prototype Release 1 Development & Technical Notes*, Michael & Susan Dell Foundation, 2010 can be found at www.texasstudentdatasystem.org/reference-docs/pdf/Public_-_TSDS_-_Lessons_Learned_Prototype_Release_1.pdf.

Similar documents will be developed to reflect the experiences and observations from future phases of the project.

Phase 2 Objectives

Having proved the feasibility of the overall approach in Phase 1, the objective of Phase 2 was to expand and extend the scope in several ways:

- Expand the scope of the dashboards to include middle school and elementary school student metrics, extending metrics that apply across the grade levels and developing new metrics where required.
- Extend the scope of the metrics to include teacher metrics and benchmark assessments.
- Add additional data sources, specifically an additional district's SIS from a different vendor, and a benchmark assessment tool.
- Refine the dashboard user interface design with new emphasis on drilldowns and the teacher's classroom view.
- Refine the prototype architecture to consider future large-scale deployment issues.
- Expand educator feedback and exposure to an additional district and a regional Educational Service Center (ESC).
- Increase the depth of user feedback and interaction to include a hands-on workshop.

Executive Summary

The Phase 2 effort further validated the major conclusions of Phase 1, as follows:

- **Local education agencies (LEA) have plenty of data to drive relevant dashboard metrics.** After looking at two districts and collaborating with a regional ESC, the effort proved the availability of data to drive actionable metrics. As discussed later in this document, differences in the data from different sources, sometimes subtle, do pose a challenge.
- **The CDM reference model and associated XML standard provide the backbone for flowing LEA data end to end through a statewide data system.** The expansion of scope in Phase 2 extended the CDM in predictable ways as new dashboard metrics were defined and new types of data were incorporated. But while differences in data from different sources sometimes posed challenges in interpreting, parsing, and mapping, the CDM required no major structural changes.
- **The internal DCD system design conceived during Phase 1 continued to prove out in Phase 2 without major changes.** Phase 2 began the investigation of several larger system deployment architecture issues.

Since Phase 1 proved the basic underlying concepts and architectures, the project team largely assumed that the second phase would be more predictable and production oriented. However, Phase 2 did present some difficult challenges, leading to a deeper understanding of large-scale, statewide dashboard development.

The major conclusions are summarized below. A more extensive discussion follows this executive summary.

Data anomalies are incrementally discovered as real data are used to calculate the metrics. A *data anomaly* is described broadly as any characteristic of the source data that results in the dashboard metric value appearing to be skewed, erroneous, or counterintuitive when it may be otherwise correct.[4] Even though most of the metrics in Phase 2 had already been developed and refined in Phase 1, this phase demonstrated that new data sources will uncover new data anomalies. The process of discovering and resolving data anomalies occurs iteratively - as sets of data anomalies are resolved, new and different classes of data anomalies present themselves.[5]

A vendor's choice in its underlying database along with its structures and architecture may increase extraction efforts. Although the team had anticipated encountering a variety of source data systems presenting varying levels of difficulty in mapping, the effort to map the SIS data from the Allen Independent School District (ISD) took longer than anticipated. The district's SIS product uses the Progress database, for which up-to-date drivers were not available for the Extract, Transform, and Load (ETL) tool. In addition, the product frequently embeds arrays in single fields, requiring additional ETL development to parse the data. Finally, the product uses a separate database for staff information, requiring special logic to map staff information in one database to teachers identified in the other. While these challenges were not unexpected, future efforts need to more closely examine the underlying database to more accurately predict required hours and resources for the mapping process.

Differences in districts' use of the same tool affect extraction efforts. Based on district and state education agency experience, the team anticipated significant variability in the way districts configure, implement, and use tools. Simply having the same tool does not guarantee similar implementation or usage. This was the case with Lubbock and Allen ISD's usage of Eduphoria, an assessment management tool. Many different types of assessments (e.g., state assessment practice tests, benchmark assessments, six-week exams, quizzes, etc.) are captured and managed within the tool. These assessments are organized by users into folders and subfolders. The process of identifying the relevant benchmark tests, their subjects, and their grades was easier for one district because of its standard conventions for naming assessments and organizing folders.

[4] For example, a middle school student's transcript showed only one course in a grade level; further investigation indicated the student was promoted two grade levels in one year, attributing only one course to the lower grade level.

[5] In the later section on Data Anomalies, a detailed example is presented that shows how the Algebra I metric was refined five times in different *waves* of discovery.

A continuous process of stakeholder engagement is key to realizing readily understandable and actionable metrics. Throughout Phases 1 and 2, stakeholders were engaged at every opportunity to provide feedback on metric requirements, concepts, wireframes, and prototype demos. This feedback was used to make hundreds of adjustments and refinements to the metric definitions, their underlying data mappings, and the dashboard user interface. The results were evident in a hands-on workshop with educators, as summarized in the next conclusion.

Dashboard measurements evolve through a *metric refinement cycle*. In Phase 1, the team identified a set of indicators predictive of student performance from the body of academic research and best practices that were responsive to an initial set of requirements elicited from stakeholders. Metrics were designed that could be derived from existing data at the campus and district levels. Metric visualizations were validated and refined in a set of regional workshops involving more than 2,600 educators. The metrics continued to be refined as they moved from concept to implementation and as they were exposed to live student data. User feedback on the operational dashboards continued to refine the metrics and their visualizations. Those looking to leverage the TSDS metrics[6] can shorten their time to deployment by taking advantage of the research, development, and refinement activities accomplished by the project.

In the first hands-on trial use of the dashboards, educators were able to quickly comprehend the meaning of the dashboards and formulate actions. At Lubbock ISD, 84 high school teachers, department heads, principals, assistant principals, and counselors, as well as some district administrators, were provided hands-on use of the prototype student, teacher, and campus dashboards loaded with their live student data. Without any formal training and with little more than a quick demo, the users were able to understand the dashboard visualizations, comprehend the information that the dashboards presented, and draw conclusions. In a short period of time, educators collaborated to plan new campus-level initiatives, singled out students for specific interventions, and identified areas to improve the quality of data in their source systems.

[6] The TSDS metric documentation can be found at http://www.districtconnections.com/Dashboards.htm.

Building Upon Phase 1 Process Lessons

As reported in the *Phase 1 Lessons Learned* document, many things went well, and some did not. In Phase 2, the team sought to improve the process by maintaining the successful aspects and making adjustments where indicated.

The project retained the same multifaceted project team but expanded for additional ETL development, test, documentation, and metric research capacity. The multistep ETL architecture was retained, with the process being transferred to three new ETL developers. New lead roles evolved in the team to mentor and direct the new ETL and documentation resources. Additional research staff were also added to develop the middle school and elementary school metrics.

The Test-Driven Development (TDD) approach was continued with additional investments in the underlying automated build and test infrastructure. The agile development process was tuned to include more detailed planning while specifically allowing time for changes in anticipation of a more production-oriented Phase 2. The test staff was specifically directed to focus on data anomalies.

The project continued to seek a high level of stakeholder feedback and to follow through with that feedback to refine the metrics and improve the user interface.

The collaborative process of defining and refining new metrics and involving researchers, analysts, and developers was continued and expanded. The metric researchers would hand off new metric definitions in written and verbal “brain dumps” to the analysts. The analysts would perform the data mapping and business rule documentation and prepare a graphic mockup for approval. During this process the developers would query the actual district data to verify that the desired result would be achieved. Once finalized, the analysts would prepare a printed metric package and brief the developers.

What went well

The team processes and dynamics continued to work well. The new ETL developers quickly learned the project’s TDD approach and were productive. The additional metrics researcher, an educator with more than 10 years of classroom experience, proved to be a valuable addition.

The collaborative metrics definition process proved to be effective and productive in driving to precise metric definitions and visualizations.

The focus on stakeholder feedback was key in refining the metrics and the prototype as evidenced in a positive hands-on workshop at Lubbock ISD, as discussed later in the document.

What went not so well

Several tasks were underestimated, though for different reasons:

- The high level of user feedback, while positive for the project, resulted in more changes to the original high school metrics than anticipated. In the first month of Phase 2, more than a hundred modifications and perfections were made, primarily to tweak the business rules or refine the visualizations, all without changing the original notion of the metrics.
- Mapping the Lubbock ISD data in Phase 1 was easier than expected given the district's use of a relatively modern SIS and central data governance. Mapping the Allen ISD data was more difficult due to an older database underlying the SIS. While the issues were not unexpected, as discussed later in the document, the effort took longer than anticipated.
- Data anomalies, while expected, were discovered at a rate and characteristic higher than anticipated. Unlike software bugs, which generally follow a trend of discovering a large number initially and then tapering off, data anomalies present themselves in waves. As sets of data anomalies were resolved, new and different classes of data anomalies presented themselves. For example, the metric for whether a student has taken and passed Algebra I by an appropriate point in his or her education was refined five times due to various data anomalies.

Phase 1 was challenged by many *firsts*: first set of metric definitions and visualizations, first application of the CDM, first data extractions and mappings of district data, first dashboard implementations. Because Phase 2 was merely expanding upon those Phase 1 efforts, the development team hypothesized that Phase 2 was going to be less discovery and more production. This proved to be incorrect as the number of refinements and data anomalies was larger than expected. Early attempts to manage and limit refinements to meet the production schedule were not appropriate. Midway through the phase, the team adjusted to a more agile process, controlling scope through weekly prioritizations.

Lessons to carry forward

The importance of prototyping and piloting activities was reinforced as significant learning and discovery continued during Phase 2. Sufficient time needs to be allocated for iterative change and refinement.

The high level of stakeholder feedback is worth the additional efforts associated with the refinements.

Refining and Expanding the Metrics

At the end of the first phase, the project held a set of feedback sessions with Lubbock ISD (the first district to supply data for the prototype) and with other educator groups, both in person and through webinars. These stakeholder sessions resulted in a number of suggestions for improvements and refinements to the original high school student and campus dashboards. These changes can be categorized as follows:

- Refinements to the user interface (e.g., showing the counts for both numerator and denominator for campus metric percentages or showing deltas in meeting campus percentage objectives as absolute values);
- Refinements to metrics (e.g., modifying Algebra I mastery to reflect students taking the course in middle school);
- Suggestions for inclusion of additional metrics (e.g., early language assessment metric); and
- Resolution of bugs in the dashboard software (e.g., state assessment objectives drilldown not appearing for some students).

The number of student and campus metrics grew 65 percent during Phase 2, from 34 metrics in Phase 1 to 56 metrics in Phase 2. This growth was due to following reasons:

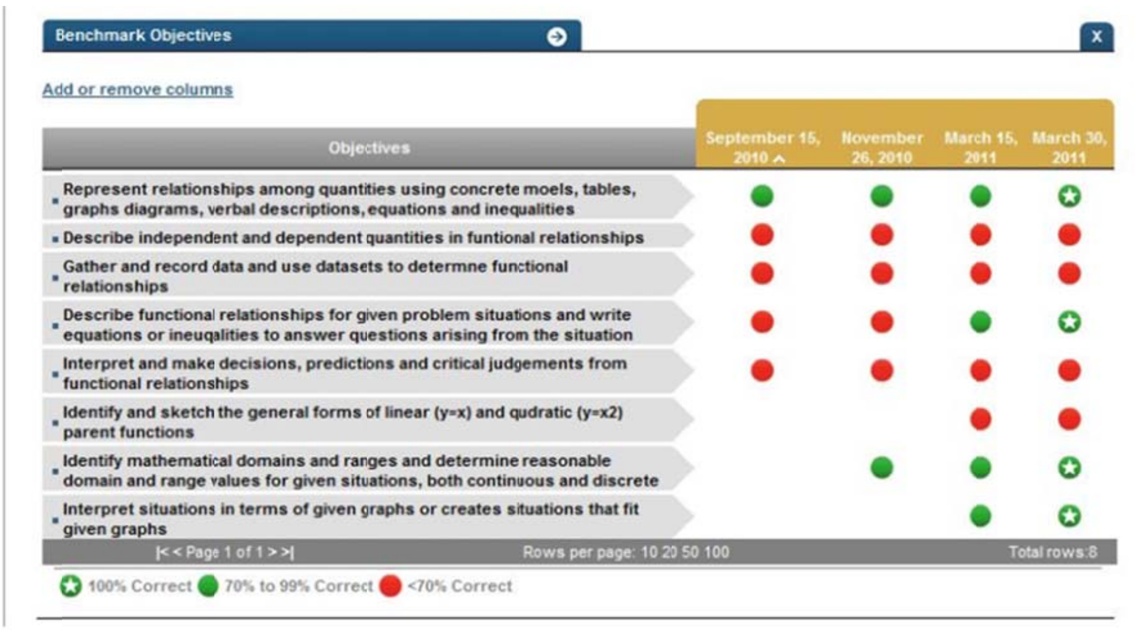
- Expansion of the dashboards from high school to middle and elementary schools;
- Addition of benchmark assessment metrics;
- Addition of teacher metrics at the campus level; and
- Addition of longitudinal campus metrics for graduation, completion, and dropout rates.

The middle school and elementary school dashboards were largely an evolution of the high school dashboards. Many of the high school metrics were adapted to middle and elementary school, specifically those for attendance, discipline, grades, and state assessment scores. Adapting some metrics (e.g., grades and attendance) required more significant changes for elementary school. The student information and overview pages required only minor changes. New metrics to support other district assessments (such as reading and benchmark tests) were the most significant additions.

Both Lubbock and Allen ISD use a tool called Eduphoria to manage their assessments, importing their state assessment scores (Texas Assessment of Knowledge and Skills or TAKS) and managing the creation and scoring of additional tests, including TAKS preparation tests, benchmark tests, class tests, and quizzes. Eduphoria maps each question and individual student results to the specific learning objectives of the curriculum. For Texas public schools, these learning objectives are defined as *student expectations* in the Texas Essential Knowledge and Skills (TEKS).

Based upon past experience, the research team anticipated that benchmark tests would be administered at the beginning, middle, and end of the year and would test all of the grade's or subject's required student expectations each time. Such an approach would lend itself well to a *test-centric* visualization, showing the student's overall score progression over time. However, neither Lubbock nor Allen ISD chose to administer their benchmark tests using this structure.

The team found that the Lubbock and Allen campuses had the freedom to use benchmark tests to address their particular needs and instructional style. As a result, there was little consistency across campuses in the timing or content of benchmark tests. Some student expectations were tested only once, some were tested multiple times, and some were not tested at all in benchmark tests. This variation led the research team to create an *objectives-centric* visualization, showing the student's results for each student expectation across multiple tests, as shown below. This visualization is meaningful independent of the number of times a student expectation is tested.

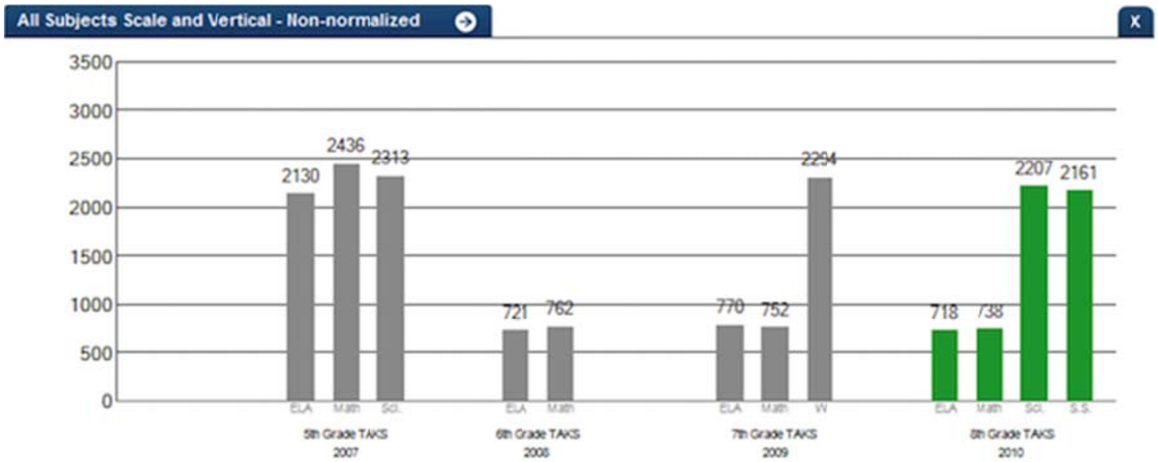


TEKS Objectives-Based Benchmark Metrics Drilldown

Starting in the 2011–12 school year, the current statewide TAKS tests will be phased out and replaced with a new STAAR (State of Texas Assessments for Academic Readiness) program, and the high school subject-based TAKS tests will be replaced with a set of end-of-course assessments. While the details of the program are still being defined, TEA has announced that the new STAAR tests will have different scoring that is similar but not comparable to the current TAKS scores.

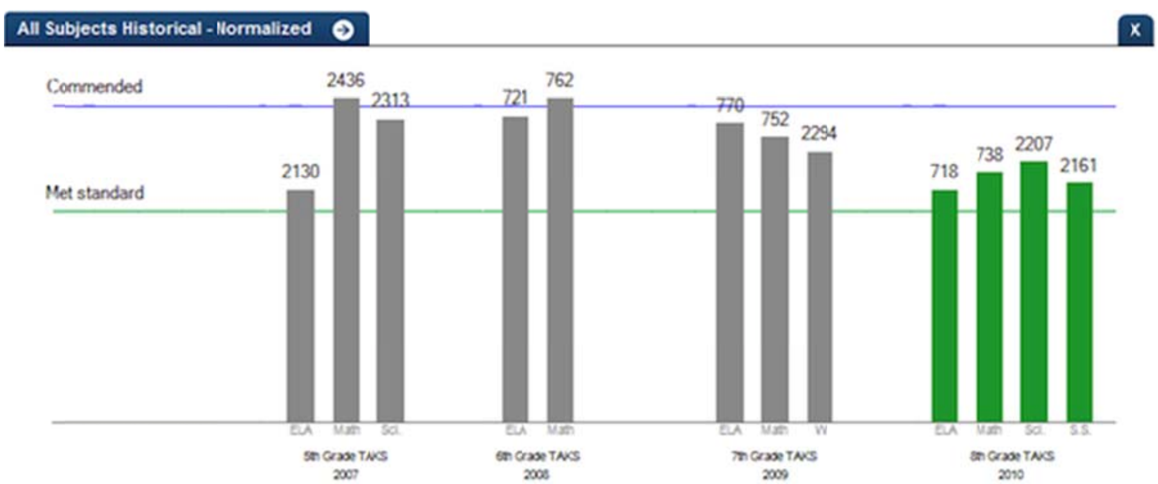
In response to these future changes, the team revisited the Phase 1 TAKS metrics and visualizations. While the dashboard's red/green determination of whether a student met the minimum standard (or similar) would still apply, the historical drilldowns would be a problem because of the different scoring systems.

For example, the following illustrates the problem when displaying scale scores (the scores in 1000s) in the same graph as vertical scores (measured in 100s).



Non-Normalized Historical TAKS Drilldown

In response, the team investigated a number of different visualizations and settled on normalized view in which the graph is computed based upon raw scores that show the cut scores for the “met standard” and “commended” levels, as shown below. The associated scale or vertical score is displayed for information purposes.



Normalized Historical TAKS Score Drilldown

What went well

The process of expanding the dashboards from high school to middle and elementary school was accomplished efficiently in the time estimated, following a methodical and repeatable process.

Both the benchmark tests and the TAKS-to-STAAR changes posed unique visualization challenges to the team. In both cases, new visualizations were discovered by iterating through numerous alternatives.

What went not so well

The number and scope of changes to the high school metrics exceeded estimates, including more than 100 specific issues and taking approximately six weeks to resolve. This additional effort, however, later resulted in positive user reviews.

Lessons to carry forward

Certain metrics (as was found for benchmark assessment metrics) may need to be fine-tuned for LEA policies and usage conventions. One cannot expect all metrics and visualizations to apply to all education organizations.

Dashboard metrics will need to be continually updated and refreshed over time to reflect changes in education policy (as in the TAKS-to-STAAR transition).

Mapping New Data Sources

Phase 2 added the mapping of SIS data from a second district — Allen ISD — as well as the benchmark assessment data from a tool used by both Allen and Lubbock ISD. To assist in answering mapping questions, the project enlisted the help of the ESC from region 10. ESC 10 had developed Empower, a data warehouse with analytical display offerings for subscribing districts, and thus had experience mapping data from particular source systems in use at Allen ISD.

The mapping of data from the Allen SIS took longer than anticipated for a number of reasons, as described below.

With Lubbock's SIS, the team was able to directly extract data from the underlying database using a connection between the database and the ETL tool. The Allen SIS uses the Progress database instead of one of the market's more common database management systems.[7] The ETL tool did not have drivers for the Progress database, and generic SQL drivers did not work properly. Attempts to find and install an up-to-date driver failed. A workaround was ultimately used, dumping the Progress database tables into an intermediary SQL server database with the necessary drivers for the ETL tool.

[7] The most popular enterprise database management systems are Microsoft SQL server, Oracle, IBM DB2, and the open-source MySQL.

The Allen SIS database was optimized in many places to store an array of values into a single database field. For example, the following convention was used to store class period attendance in a single field:

- **;;E;E;E;** meant that a student had excused absences for class periods 3, 4, and 5 but not for periods 1, 2, or 6.
- **;T;;;;;T** meant that a student was tardy in periods 2 and 8.

Each time such a convention was encountered, the ETL developer was required to write special logic to parse and interpret the data.

Another challenge in the mapping of Allen SIS data occurred in the processing of staff data to support teacher dashboard metrics. Teachers are associated with campuses and course sections in the student SIS database. Information about the teachers — for example, their years of experience or their certifications — is stored in the human resources SIS database. Merging these data required special logic to match teachers in the student database with those in the human resources database.

This merge meant that mapping the Allen SIS took twice as long as mapping the Lubbock SIS. Mapping the Lubbock data during Phase 1 took approximately 400 person hours. By contrast, mapping the same Allen data took 800 person hours, with 240 person hours spent establishing a connection between the Allen Progress database and the ETL tool and 560 person hours spent mapping the data. Roughly 25 percent of the mapping time was associated with parsing arrays stored in individual fields and other Allen data variations.

Both Lubbock ISD and Allen ISD use the same assessment tool — Eduphoria — to define different types of assessment data, capture student scores, and analyze assessment data. The process of mapping the raw data from the tool was straightforward: The vendor provided an XML export format that was readily mapped to the CDM XML standard. However, the individual assessments in the tool did not have *metadata* that identified the type of assessment (e.g., TAKS, TAKS preparation, benchmark test, class subject exam, or quiz), the subject of the exam (e.g., English, mathematics, science, or social studies), or the grade level for the exam.

For Lubbock, the metadata could be determined by the standardized folder structure and naming convention the district adopted for the exams. Without such a set of standard conventions, the Allen benchmark assessment could not be identified, and the data could not be successfully mapped. (With help from district and campus personnel, the missing information could be derived, of course. However, one of the prototype goals was to require little or no data creation by participating district personnel.)

What went well

In general, the process of doing data extractions and mappings has proved to be repeatable and transferrable to new personnel.

Relationships with key personnel who understand the source data — at Lubbock and Allen ISD, ESC 10, and Eduphoria — provided invaluable assistance in deciphering source data and obtaining feedback for alternative visualizations.

What went not so well

In Phase 1, the relative ease with which the Lubbock SIS was mapped resulted in the assumption that mapping the Allen SIS would go just as well. While the problems encountered were not completely unexpected, the additional effort required was an issue throughout Phase 2.

The team did not anticipate the scale of the impact caused by differences in each district's use of Eduphoria. The experience points to the need for standardization in how local districts use the tools that exchange data with other systems. However, had the tool specifically collected and stored the required metadata and used an existing standard means of data exchange, reliance on the creation of local conventions may not be required.

Lessons to carry forward

More initial analysis of data sources is indicated to identify potential problems and issues in the tools underlying the database and in local conventions in districts' use of the tools:

- Examine the underlying database technology and its ability to support extracts.
- Examine the database schema, how it is organized, and how the table primary keys are constructed.
- Conduct a preliminary coverage analysis for each of the interchange schemas, examining any existing views that extract similar data.
- Browse the underlying data and enumeration code values looking for areas where additional transformation logic may be required.
- Identify areas where student and/or staff identity mapping will be required.
- Identify areas of local extension, for example local course codes, local discipline codes, etc.
- Investigate the number of years of available historical data and if there are any format or structure changes for those years.
- Identify areas where campuses differ, for example in bell schedules or use of benchmark assessments.
- Identify what special accommodations are provided to collect accountability submission data and allow entry of special codes.
- Interview data administrators and report writers for known idiosyncrasies.

The need to pilot dashboards using a thorough cross-section of data sources is reaffirmed.

For the future TSDS deployment, vendor and user relationships will be critical to create and maintain ETL programs for the number and variety of source systems. Vendors can further ease both the one-time transition and ongoing maintenance of data exchange by building standard TSDS data interchanges into their systems. In any case, data system vendors will need to inform districts of underlying changes in structure or semantics that may affect loading or inhibit the ability to have comparable historical data.

Data Anomalies

Phase 2 expanded the metrics in several dimensions: new grade levels, new metrics, and new metric data sources. As expected, this expansion provided the opportunity for discovering new data anomalies. Each data anomaly that is discovered and explained or resolved better prepares the dashboards for statewide rollout.

In this paper, a *data anomaly* means any characteristic of the source data that in any way results in the dashboard metric display value to be skewed, erroneous, or counterintuitive when it may be otherwise correct.

To achieve data quality in any rollout effort, resolving data anomalies is time consuming because they can have a number of root causes, including:

- Missing data
    - The data source is missing certain types of data for all or most students.
    - The data source is missing certain types of data for select students.
    - Student data are missing due to transfer(s) between schools.
    - Data are missing because the district has not yet entered or loaded the data.
    - The district has not loaded, or has only partially loaded, historical data into its source systems.
- Erroneous data
    - The source data for a specific student are erroneous from the source system.
    - The source data for all or a segment of students are erroneous.
    - Student data are erroneous due to an identity-matching problem.
- Inconsistencies that may ultimately end up being either erroneous or accurate
    - The source data for a student are inconsistent with other data about that student.
    - The source data are correct for a student but do not appear consistent with other students' data.
    - Changing student populations due to transfers and dropouts make aggregate metrics appear to be inconsistent.
    - Different aggregate totals for one metric are not consistent with totals for other metrics for the same campus.
    - Aggregates or percentages are not consistent with other published metrics (e.g., accountability metrics) due to timing or differences in cohorts.
    - Aggregate metrics are not consistent across campuses or districts due to different policies, practices, or conventions.
- ETL or programming errors

    - The source data are collected and/or stored differently in different grades.
    - The semantics (meaning) for a type of data are misinterpreted by an ETL programmer in computing a metric.
    - The ETL program or the metrics display logic is erroneous.
    - The ETL program should be processing new or different data.
- Local differences
    - Different data sources (from different tools and/or from different districts) store or represent the same data inconsistently (e.g., having different local codes).
    - There are local process or business rules that must be taken into account (e.g., different grading periods or different grading schemes).

The following are a set of example data anomalies discovered in Phase 2 that illustrate the diversity of data anomalies.

Historical data were limited in the Allen SIS. Historical student data were difficult to obtain for Allen ISD for a number of reasons: (1) The district changed to the current SIS two years previously; (2) the SIS does not have structures to store historical transcript information apart from current course grades; and (3) the district archives its student data each year and initializes a clean SIS database without historical information. Therefore, some high school metrics could not be correctly populated. The initial data analysis of district sources needs to investigate the availability of historical data.

Drastically different results for the campus metric for Repeating Failed Courses revealed different business processes between Lubbock ISD and Allen ISD. The high school metric for Repeating Failed Courses measures the percentage of students at a campus who are currently retaking a course that they previously failed. The team questioned the two districts about the fall results: Allen high schools had no students repeating a failed course, while Lubbock high schools had an expected percentage. The team discovered that Allen ISD has a policy that requires students who fail a course to repeat it in a summer semester, thus explaining why no students were repeating a course in the fall semester. Lubbock ISD students repeat failed courses during the regular fall and spring semesters of high school. As in this case, data inconsistencies may not be erroneous but may accurately reflect the result of local processes or conventions. This example highlights the need for the indicators in the TSDS system to acknowledge local variation (e.g., by enabling multiple ways of measuring a single notional indicator and allowing district and campus personnel to select the measures that are meaningfully aligned with local practices).

Computing Credit Accumulation for Allen required a deep semantic understanding of the district's enumeration codes. Many source systems constrain data values to choose from a set of enumeration codes that are specific to the SIS and may be specific to the district. Understanding these enumeration codes is critical to correctly interpreting the data. For example, the initial Allen mapping pulled student credits from a transaction table, pulling credits awarded in semester 1 (S1), semester 2 (S2), and yearly (YR). The resulting metrics showed a large number of students that appeared to be missing credits. Discussions with Allen indicated that there was another code (YN) for year-long courses with no semester exams. Correct extraction requires a deep understanding of internal enumeration codes and local conventions. Implementing the TSDS solution statewide will require managing the balance between enabling local conventions that correctly model how districts actually operate and ensuring that local conventions are correctly and exactly translated and standardized by the statewide platform.

State Assessment (TAKS) Nonparticipation metric computation had errors due to the complexity of the dataset. The TAKS score reports contain a code for students who registered to take the test but did not for any number of reasons (e.g., absent, exempt, parental waiver). A data anomaly was discovered when the number of students with results plus those reported as not participating was not consistent with total enrollment. Two root causes were found: The program was not processing the results of the TAKS test modified for students with disabilities, and one of the reasons for nonparticipation, that the student previously met standard, was not interpreted correctly. Complex data sources such as the TAKS results are the most prone to ETL errors.

Allen ISD had a large number of discipline incidents. The discipline metric showed 56 percent of Allen ISD students had one or more discipline incidents. Further investigation indicated that the metric calculation was correct based on the underlying data but that the vast majority of incidents were "violation of school of conduct," which could be as minor as chewing gum or inappropriate dress. A future enhancement is planned to extract and display a secondary reason for these more minor incidents. As was the case here, local conventions for recording data may be the source for what appears to be inconsistent.

Early childhood grades were missing for Allen ISD. Course grade data for kindergarten, first, and second grade students were only available for noncore subject areas such as art, music, or physical education. Inquiries to Allen ISD indicated that for these early grade levels, core subject grades are not logged each period and, in some cases, not logged until after the end of the school year. These early grade levels have objectives-based report cards. Local decisions for when to enter or load data will affect dashboard data.

Computing daily attendance varies by grade level. For high school and middle school, daily attendance is computed based a *homeroom* section or other designated section during the day (e.g., second period). Elementary schools may not have designated sections (other than for special subjects such as art, music, or physical education) and therefore record daily attendance in different tables in their SIS database. Thus, the same data element, in this case daily attendance, may be stored differently in the same source system for different circumstances, such as grade level.

Computing class period absence rate is complicated by different bell schedules. Class period attendance is reported by exception — an absence is recorded but being present is not. Computing the class period absence rate means dividing the number of class absences by the total number of class periods. The initial computation assumed that students attended each class period every instructional day. However, many high schools and middle schools have bell schedules where class sections may not meet on every instructional day. This required the CDM to add a structure to reflect the bell schedules for each campus and for the class period absence rate to be adjusted accordingly. Future refinements to the CDM are expected where there are wide variances in district processes.

When student or staff data are stored in different data sources, identity-matching problems may occur. In Skyward, Allen ISD's SIS, teacher data are stored in two databases. Teacher assignments to sections are stored in the student module. Teacher experience and certifications are stored in the human resources module. Matching teachers across the databases identified the anomaly of two teacher records with the same Social Security number and different names. Identity matching, when required, is a critical part of the extraction process.

As discussed previously, discovering data anomalies is an iterative process. An example of this iterative nature is found in examining the history of the Algebra I metrics. Research found that a student's success with Algebra I was an indicator of college readiness. Two high school metrics for Algebra I were defined in Phase 1: whether the student has/is taking Algebra I and whether the student passed/is passing the course. From their initial definition, the Algebra I metrics have incrementally evolved through many changes, as follows:

1. Initially, the numbers of students who had taken or were taking Algebra I in Lubbock high schools were noted to be unusually low. Investigations indicated that there were several local course codes for variants of Algebra I that were not known. The adjustment was made to consider all of the local course codes for all variants of Algebra I for Lubbock high schools.
2. The data anomaly was discovered that students who were currently taking higher-level mathematics such as Geometry and Algebra II had metrics indicating they had not previously taken Algebra I. The cause of this anomaly was that the several students were missing their transcript information either because they were transfer students with an incomplete transcript or they took Algebra I in middle school and those records were not available to the system. The metric definition was updated, for those transcripts that do not show Algebra I, to infer that Algebra I was previously taken and passed if the student was taking higher-level mathematics courses.
3. Through additional discussions with Lubbock curriculum specialists, it was discovered that if a student did not take Algebra I in middle school, the student was required to take it in 9^{th} grade. Thus, if a student was taking Algebra I in 10^{th} grade, the metric showed positive, even though the student was lagging in his or her mathematics education. The metric definition was updated to measure whether the student is/had taken Algebra I in the 9^{th} grade.
4. At Lubbock, individual grades and credits are earned in each semester, and a final grade is not assigned. Thus, if a student failed first semester and passed second semester, the

system could score the passing metric positively. The ETL process was updated to average the two grades to determine whether the student passed.

5. In the Allen student information, many students have transcripts that record credits earned, but without a recorded grade. The ETL process was updated to alternatively indicate whether a student has taken and passed Algebra I based upon credits earned.

Key members of the project team and representatives from TEA and MSDF took part in a Joint Application Design (JAD) session. The session considered the following major sources of data anomalies:

- **If certain data are not loaded at an *expected* frequency and latency, the metrics may reflect old or incomplete data.** For example, if the latest grading period grades have not yet been loaded, the list of students flagged for poor classroom performance may no longer be accurate.
- **Late student enrollments result in data anomalies from having incomplete data**. For example, absence rates would look lower if not compensated for late enrollments.
- **Missing or incomplete transcripts affect many student metrics**. For example, missing transcripts would affect most historical drilldowns, as well as metrics requiring transcript information such as Credit Accumulation.
- **Timing of TAKS test results availability causes specific issues with viewing results.** During most of the school year, the last version of TAKS results is all that is available for a student. Late in the school year, TAKS is administered over different dates for different subjects and grades and with different administrations of the same test (for makeups). Because administrators want to see the current year results as they become available, the dashboards must deal with different years' assessment results.

The JAD session resulted in a number of specific recommendations to change the labeling of metrics or to add footnotes to properly inform the educator — most of which were implemented or demonstrated during Phase 2. In addition, the group recommended future phases investigate developing special programs and/or reports to identify data anomalies proactively and alert district administrators when expected data have not yet been loaded.

What went well

The project team diligently discovered and analyzed data anomalies and, when necessary, modified the metric definition and/or the ETL programs appropriately. The DCD has been developed to function with incomplete or partial data. The team's hypothesis is that as the data become more accessible and visible to educators, incomplete or incorrect records will be addressed, thus improving data quality.

What went not so well

The team was too optimistic that experience with one district in Phase 1 would greatly reduce the impact of data anomalies discovered in Phase 2. The project underestimated the impact of data anomalies in three main areas:

- The project assumed that most of the data anomalies were already discovered for the high school metrics developed in Phase 1.
- The project underestimated the new data anomalies that would result when metrics developed for high school were expanded to cover middle and elementary schools.
- Differences in district processes and conventions that affect how data are stored were larger than expected.

Lessons to carry forward

Discovering and resolving data anomalies are a natural part of the refinement cycle for dashboard metrics. However, such data quality activities will remain time consuming and need to be specifically planned for in future rollout efforts.

The mapping of new data sources should be preceded by an extensive examination of the underlying data store; the data themselves; and the processes, practices, and conventions of the data's use.

Future dashboard development effort will investigate building specific tools and reports to allow districts to identify and resolve data anomalies both during initial deployment and in ongoing operation.

Refining the Data Model

The project continued to track the development of national education standardization efforts, including National Education Data Model (NEDM),[8] Common Data Standards (CDS),[9] and Schools Interoperability Framework (SIF).[10] Both NEDM and CDS provide higher-level *starting points* for education data interchange efforts but lack the concrete details for interchange. By contrast, the CDM reference model and the associated XML standard provide the necessary detail and specificity to be the standard for submission of data to the DCD and for district data interchange in general.

SIF addresses a problem fundamentally different from that of the DCD — connecting different district and school tools to synchronize selective, shared operational data. By contrast, the CDM supports the exchange of larger, cohesive sets of data within and across organizations that may not have connected data systems.

During Phase 2, the CDM changes were motivated by the following activities:

- Definition of new metrics — for example, for staff metrics;
- Refinements discovered from extracting district data;

---

[8] National Education Data Model 2.0 can be found at http://nces.sifinfo.org/datamodel/.
[9] Version 1.0 of the Common Data Standards can be found at www.commondatastandards.org.
[10] Version 2.4 of the Schools Interoperability Framework Implementation Specification can be found at http://specification.sifinfo.org/.

- Independent review of the CDM by ESC 10; and
- Analysis of data requirements for the principal dashboards.

The Phase 2 CDM changes were evolutionary and did not require any major structural changes. The changes are described in detail in *Canonical Data Model (CDM) Phase 2 Changes and Rationale*.[11] A representative overview of CDM changes is listed below:

- The Staff association to a District indicating work assignment was generalized to denote assignments at any level of organization (i.e., ESC, district, or campus).
- The Student Expectation attributes were reworked upon the actual metadata for the TEKS.
- A new attribute was added to Test Assessment to support benchmark tests spanning multiple grade levels.
- A new entity called Academic Week was added to support future analysis based upon experience from ESC 10.
- The Course entity was reworked to denote courses offered by a campus and identified with a local course code. State course code is now a descriptive rather than a key attribute.
- New Leave Event and Certificate entities were added with associations to Staff to support new metrics.
- A new Bell Schedule entity was added to accurately compute and represent class period attendance.
- A new Cohort entity was added with its association to Student to track cohorts in the principal dashboard.
- Attributes were added to support Accountability Ratings for Education Organizations for the principal dashboard.
- A new Post-Secondary Event entity and its association to Student were added to support the principal dashboard.

It was expected that variances in a second district's data would result in more substantial CDM changes. In reality, most changes came from consideration of new metrics (i.e., expanding the problem domain) and from deeper understanding of the data (e.g., additional relationships among information).

The CDM began formal versioning and documentation of changes as it gained wider exposure within Texas and nationally. ESC 10 independently applied the CDM to load its data warehouse for the Reveal Dropout Early Warning System. The Delaware Department of Education indicated its intent to apply the CDM approach for its future statewide data warehouse.

[11] The *Canonical Data Model (CDM) Phase 2 Changes and Rationale* document can be found at www.districtconnections.com/CDMDownloads/Public - TSDS - CDM Changes and Rationale - Phase 2.pdf.

Looking toward future production requirements, an initial canonical service envelope (CSE) specification[12] was drafted to wrap data with a *transaction envelope* consisting of header information to specify the service(s) to be invoked as part of a data interchange. The intent was not to supplant the service specification mechanisms of certain technologies, such as web services,[13] but to fill the gap when using certain transport technologies widely used for data interchange, such as FTP[14] or simple dropping of a file into a shared folder. An implementation demonstrating the use of the CSE was prototyped illustrating its use in the file drop scenario.

What went well

The CDM reference model and XML schema continued to serve the project's needs in moving data into the DCD from different data sources. The assessment area continued to have significant payoff, as many different types of assessment (TAKS, SAT, ACT, Advanced Placement, benchmark) may be interchanged using the same format.

The CDM changes in Phase 2 were both small and predictable as the scope of the metrics and their data requirements expanded. The impact from additional data sources was less than anticipated.

What went not so well

The CSE was developed based on hypothetical use cases because the eventual large-scale architecture of the statewide implementation is not yet known. In the prototype, the development team initially used other conventions, such as file naming or folder structures, to avoid needing a service envelope.

This was in contrast to the concrete DCD use cases that guided the initial development of the CDM. As a result, the initial draft of the CSE is less fully "baked" than other CDM components. Additional refinements are planned in future phases.

Lessons to carry forward

The area of education data standards continue to evolve. The project will continue to track and monitor national standards activities.

Assessment vendors should look to adopt a common interchange. Having been applied in the interchange of a variety of assessments, the CDM could provide the starting point.

---

[12] The initial draft of the *Canonical Service Envelope (CSE) XML Technical Implementation Guide* can be found at www.districtconnections.com/CDMDownloads/Public%20-%20TSDS%20-%20CSE%20Technical%20Implementation%20Guide%20-%20Phase%202.pdf.

[13] The W3C standard for web services may be found at www.w3.org/standards/webofservices.

[14] The IETF standard for the File Transfer Protocol (FTP) is RFC 959 and can be found at http://tools.ietf.org/html/rfc959.

Improving the Metrics Documentation

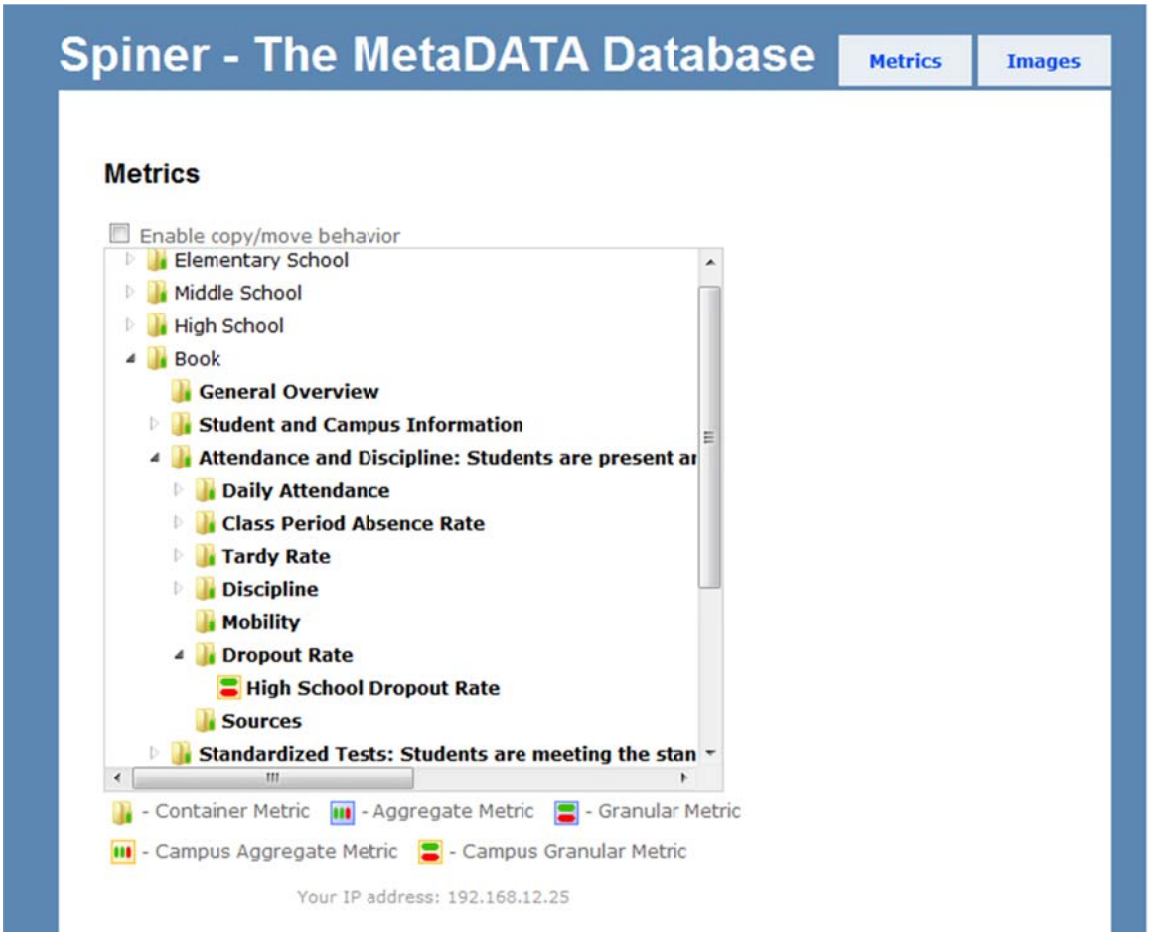
During Phase 1, a metadata database[15] and web application were built to hold and manage the metrics documentation. The metrics documentation was published for public review and use.[16] In Phase 2, the metrics documentation was reorganized and improved as follows:

- The metrics were organized into chapters. With the research basis for each metric included in the metric documentation, there was a great deal of duplication. In the new organization, the research basis and use cases are part of the chapter descriptions. The technical metric documentation is accessed from the chapters through links. Thus, the chapter descriptions serve as a bridge between educators and the more technical audience. The research explains why these particular metrics are valuable, answering the question "why would we want/need this?" The technical documentation answers "how do we do this?"
- The technical documentation was reorganized to have more tables and less prose, striking a better balance between specificity and readability.

[15] The metadata database is named "Spiner" in honor of Brent Spiner, who played Data in the popular TV show *Star Trek: The Next Generation.*

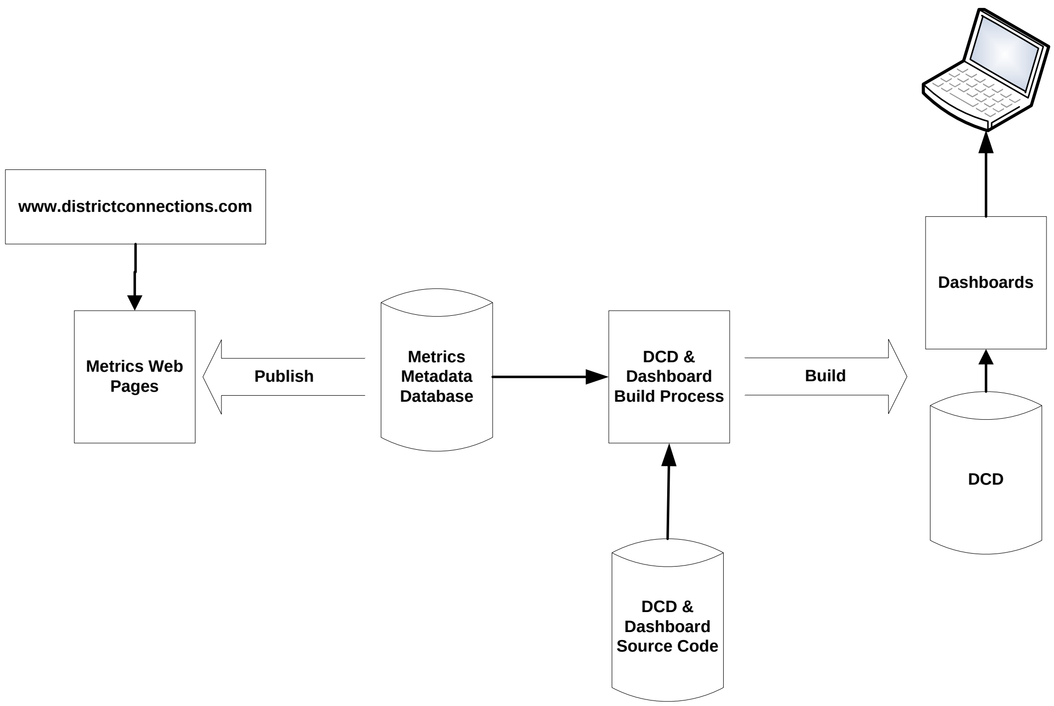
[16] The metric documentation may be found at www.districtconnections.com/Dashboards.htm.

These changes required the metadata database to be upgraded, adding a hierarchical view of chapters and metrics, as shown below.



Example Metadata Database Hierarchical Metrics View

In addition, the software build process was directly linked to the metadata database to pull the metric names, description, and dashboard page organization directly from the same source as the documentation, as shown in the diagram below.



What went well

Deriving the metric documentation and the operational organization of the dashboards ensures consistency between the two. In addition, it provides a framework that, in a future phase, could allow districts to customize their dashboards by selecting from a set of predefined metrics. Another suggested future enhancement is the capability for users to associate critiques and suggestions with the metrics in metadata database for later consolidation and refinement.

What went not so well

The metadata database is an internal tool that is growing in importance. The effort to maintain and upgrade the database is always somewhat at odds with core dashboard development work. The metadata database has quirks that will require fixing in future phases (e.g., proper formatting of tables when output in Word and HTML).

Lessons to carry forward

The investments in infrastructure to support the metrics documentation continue to provide good return as the number of metrics continues to expand and the metric definitions continue to evolve. Such an infrastructure is critical for tracking and implementing metric changes over time.

Stakeholder Engagements

At the end of Phase 1, the project had a working demo of high school dashboards populated with anonymized student data from Lubbock that were used to gather feedback from a wide variety of stakeholders at conferences, webinars, and one-on-one demos. For new visualizations, such as the teacher view or principal dashboard, the wireframe mockups were used to obtain early user feedback.

After each stakeholder engagement, the feedback was documented, and any subsequent actions determined. Some feedback resulted in immediate refinements. Other feedback was scheduled for later in the phase or for future phases.

On October 20, 2010, the project team facilitated a series of two-hour, hands-on workshops at each of the four Lubbock ISD high schools. In total, 84 teachers, principals, assistant principals, department heads, counselors, and district administrative staff used the dashboards for the first time.

The dashboards were populated with actual high school student data that was extracted from the SIS a few days before and stored in a secure and encrypted DCD database. The dashboards had an interface to the district's LDAP[17] system, allowing users to log in to the dashboards with their regular user name and password. The dashboard security limited access to only those students and views that the user should have access to according to Family Educational Rights and Privacy Act rules; for example, teachers were able to see only their classes and students. A contextual feedback function was developed for users to enter their feedback directly into the dashboard application.

The project team gave the workshop attendees an overview of the state's TSDS vision and the objectives of the DCD and the dashboards. After a brief guided tour of the campus, classroom, and student dashboards, the attendees used the dashboard system for approximately 45 minutes. During this time, they were able to ask questions of the facilitators, talk amongst themselves, or enter their feedback directly into the system.

The potential value of the dashboards was evident:

- Participants were fully engaged, as evidenced by more than 3,586 page views and 17,934 metrics viewed by the 84 participants, averaging about a page view a minute per person.
- There was overwhelmingly positive feedback on the dashboard user interface — specifically the intuitive design and ease of use.

[17] Lightweight Directory Access Protocol (LDAP) is defined by an IETF standard RFC 4510 that can be found at http://tools.ietf.org/html/rfc4510.

- All four campuses found data about students or their campus that surprised them and wanted to take specific actions right away.
- Most stakeholders used their limited hands-on time actually to *use* the dashboards in a manner appropriate for their role. For example, teachers examined specific students with known issues or found students in need of some type of intervention, counselors researched students who were not on track to graduate who might require adjustment to their class enrollments, and principals drilled down on specific campus metrics that showed their campus did not meet critical thresholds.
- Educators with varying levels of training and experience, including many with no formal training in data use, were taking notes on individual students and coming up with immediate plans of action. Several educators asked for printed lists to work from the following week.
- Educators were spontaneously collaborating on actions that should be taken with students. This collaboration was not only between those with similar roles, but also across discipline areas.
- Erroneous source system data were found at all four campuses. These errors were verified to be errors in the source SIS. This finding started fruitful conversations about the self-cleansing of the data that will happen when the tool is fully implemented and how Lubbock ISD could deal with these issues from a process perspective.

The workshop experience produced the following significant conclusions:

1. The dashboard user interface showcased the application, gaining almost instant user acceptance. After only a brief demo, stakeholders were able to navigate the dashboards effectively and quickly understand the metrics without further training.
2. Stakeholders were able to understand the meaning of the metrics, drawing conclusions about individual students or about the campus at large, and successfully formulate appropriate actions.

What went well

The project's focus on continuous stakeholder input resulted in a dashboard prototype that was readily usable and that received positive feedback from the Lubbock workshop.

What went not so well

Once the Lubbock workshop was finalized, the development team needed to shift from a prototype or proof-of-concept mentality to a pilot or production mindset, the primary impact being in our quality assurance processes. This transition did not occur as quickly as desired, and the overall timeline was affected as more resources than planned were diverted in preparation for the workshop.

Lessons to carry forward

Continuous stakeholder feedback is critically important for any dashboard development, both to ensure that the product is providing the intended benefit and to ensure buy-in from the broader stakeholder community.

The unique design style of the dashboards embodied in the user interface has, in stakeholder engagements to date, proved to be compelling and useful.

The Lubbock workshop confirmed that educators, when presented with easy-to-understand and meaningful dashboards, are enthusiastic about using the data and formulating action plans.

System-Level Architectural Considerations

During Phase 1, the prototype implementation focused solely on the internal architecture of the DCD and the associated ETL jobs to build and manage the data. During Phase 2, with the addition of a second district's data, system-level architectural questions and considerations arose, looking toward a future statewide deployment to 1,237 school districts and charters with 8,435 campuses, 4.8 million students, and 333,000 teachers.

While the purpose of the Phase 2 was not to address the large-scale TSDS deployment issues, it is useful to summarize some of the results of analyses and deliberations for future phases.

Scalability of the Current Approach

In the current prototype, the entire set of student data is loaded into a clean DCD every night. ETL jobs build the DCD and the associated metrics data store. The statistics for the two districts are shown in the table below:

	Lubbock ISD	**Allen ISD**
Number of current students	28,680	18,242
Size of data load – All Data Assessment Data Nonassessment Data	2.8 Gb 1.0 Gb (36%) 1.8 Gb (64%)	1.5 Gb 0.6 Gb (39%) 0.9 Gb (61%)
Compressed size of data load	55 Mb	23 Mb
Nightly time to process – All Data Assessment Data Nonassessment Data	45 minutes 16 minutes 29 minutes	28 minutes 11 minutes 17 minutes
Average data per student	69.7 Kb	80.5 Kb
Average load time per student – All Data Assessment Data Nonassessment Data	.094 seconds .033 seconds .061 seconds	.092 seconds .036 seconds .056 seconds

The processing time numbers above are based upon a single mid-range Windows-based enterprise server.[18]

In the final deployed system, assessment data certainly would be loaded only when new results are received as there would be no need to load that every night. Of the remaining nonassessment data, data would likely be loaded based upon differing periodicities.

To estimate the processing time for the entire student population's nonassessment data (i.e., the data that would likely be updated throughout the year), assume that processing time will be 0.06 seconds per student, as was true of Lubbock ISD data in this prototype. (While loading times may vary, Texas is investigating robust hardware and platform options that will, presumably, reduce the per-student processing time.)

At this rate, processing the regularly updated data for all of Texas' 4.8 million students would require 4,800 minutes of processing time. If it was desirable to perform a complete, nightly update of all Texas student data in six hours, this would require the processing equivalent of 14 mid-range Windows-based enterprise machines.

Future phases will analyze the various interchanges to derive the most optimal schedule for loading the various interchanges into the DCD. However, this early analysis indicates that at least in terms of loading data into the DCD from the CDM schema, scaling should not be a serious issue.

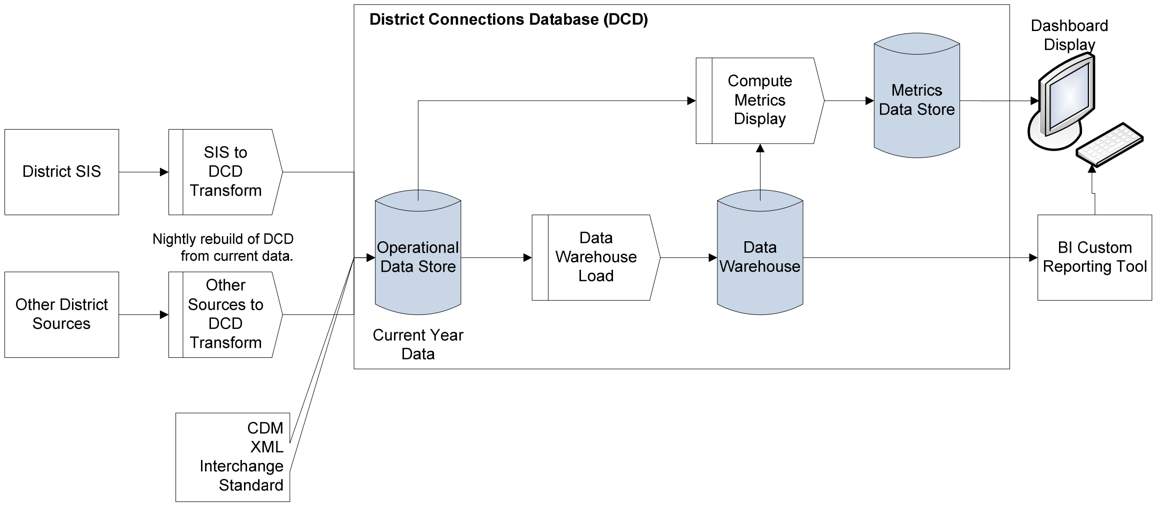
Role of DCD-Based Data Warehouse

The team facilitated JAD sessions with TEA and MSDF personnel to explore the role of a DCD-based data warehouse with supporting business intelligence (BI) analytic tools. The architectural concept is that an operational data store (ODS) in the DCD would hold the current school year's data as depicted below. A dimensionally designed data warehouse would house both selective dimensional current year data and any prior year data directly stored in the DCD. The dashboard metrics would pull data as required from both the ODS and data warehouse components of the DCD.

The data warehouse would support the dashboard metrics for the following:

- Current year student dashboard trend computation and historical drilldowns;
- Current year campus and district metric trend computation and historical drilldowns;
- Dashboard metrics with a different cohort; and
- Custom drilldowns and reporting.

[18] Specifically a 64-bit, quad core Intel Xeon 3450 (2.66 GHz) with 8 GB of memory



Architecture of DCD-Based Data Warehouse

For custom drilldowns, the dashboards would provide an interface to a commercial BI tool to support views and functions not supported by the native dashboard logic, including the following:

- Ad-hoc slicing of current year and prior year data by subpopulations;
- Selection of metrics to view;
- Selection of different graphing views;
- Historical comparisons;
- Specific cohort analysis; and
- Crosstabs.

A subset of dimensionally designed data warehouse structures was prototyped in Phase 2 to support the initial set of longitudinal dashboard metrics for graduation, completion, and dropout rates. The interface to a BI tool was also demonstrated for analyzing subpopulations for TAKS English language arts results.

Future phases should continue to build out and prove the data warehouse component of the DCD.

What went well

Facilitated JAD sessions proved useful in these early deliberations of system-level architectural considerations.

What went not so well

At this point, there is a high level of uncertainty in the strategies and architectures that will be used for statewide deployment of the DCD. Though typical of large-scale system-design projects, this resulted in many questions being deferred for future consideration.

Lessons to carry forward

The early analyses of system-level architectural issues quickly illustrated the myriad technical issues involved. Future phases will continue to analyze alternatives and use the evolving prototype to test architectural decisions.

Summary

What went well

Many of the project approaches that were used in Phase 1 continued to prove their value: continuous stakeholder feedback; continuous metric refinement based upon different data sources and stakeholder feedback; an agile, test-driven development process; and a continuously refined user interface focused on ease of use and understandability.

What went not so well

Phase 2 presented difficult challenges dealing with data anomalies and new data sources.

Lessons to carry forward

The team learns something from every instance of stakeholder feedback, from every new district, from every new data source, and from every new data anomaly. The prototyping and pilot process is critical to satisfying the long-term goals of the project for student performance improvement.