

The Need for Data Governance

Education accountability, management and targeted interventions necessitate student longitudinal data systems. Advances in technologies have made such systems possible, but being able to build such systems results in little benefit if there is no data governance. While there are various definitions of good data governance they must include some organizational structure that defines data policies and procedures along with a plan to implement them, all to ensure quality data. This brief was written by EIMAC's Longitudinal Student Data Systems Task Force following a discussion on the importance of data governance at their October, 2007 meeting.

Data Quality Drives the Need for Data Governance

Quality data for the purpose of improving student achievement is the basic objective of a longitudinal student data system. Without deliberate governance, data quality will suffer and the information used for critical decision-making will be questionable.

All state education agencies will see themselves in the following typical examples of the consequences of poor data quality: One local agency forgets to report its homebound students (**completeness**); a local agency labels all students leaving school for unknown reasons as transfers (**accuracy**); an analyst in the curriculum department uses one rule to calculate class size while the public relations department calculates class size using another rule (**consistency**); an import program does not check for a "M" or "F" in the gender data element (**validity**); two schools with the same name but from two different local agencies have their data comingled (**integrity**); an unauthorized data analyst accidentally changes every teacher's year of birth to 1900 (**security**); data is late arriving from an assessment vendor causing student profiles to be unavailable for parent-teacher conferences (**timeliness**); a chief is embarrassed at a meeting because an expected online report is not available (**accessibility**). These data characteristics must be addressed in a systematic and rigorous fashion to ensure data quality. Achieving each of these imposes technical and organizational requirements which must be governed – or states will fail in their efforts to significantly impact student achievement by making decisions with inaccurate data.

The impact of deficiencies in data quality is more obvious in some cases but all can have serious consequences. In the era of high stakes testing and accountability a mistake in one student's days of enrollment, ethnicity, grade level or test score can make the difference in a whole school or district making or not making a critical performance goal. Likewise, inaccurate student data on a transcript can make the difference in acceptance or rejection by a college.

The Components of Good Data Governance

Leadership: A state's education chief must champion data governance to ensure successful acquisition of resources, communication of goals and organizational change. A data governance board made up of senior leadership from the agency's departments should meet regularly for strategic data planning.

Data management staff includes a data quality manager or data administrator, a chief information officer (CIO), IT director and a data steward coordinator; some of these roles may be combined. These individuals must be part of the governance board which must have direct access to the chief. Data stewards, who must be skilled data analysts and content experts, are responsible to their department directors and work closely with the data quality manager and the IT staff.

The consequences of insufficient leadership cannot be minimized. Without it, a general lack of commitment will pervade the organization followed by failure to meet some, if not all, of the aspects of data quality.

Data Quality Management: The data quality manager establishes policies and practices that assure quality data. For example, student data should be derived directly and without duplication from student information systems. Validation rules, including completeness checking, must be present and enforced.

Those responsible for data governance at the state agency must provide for: adequate time for change; staff development; oversight of data use; scheduling of data delivery; and clear lines of authority. Documented data standards, calendars and exchange protocols (such as the School Interoperability Framework standards) are mandatory.

Local agencies need their own designated data managers. Each local agency head must promote and be accountable for accurate data gathering at the school level and the successful submission of data from the local agency to the state agency and other education entities. Some states mandate statewide certification programs for anyone maintaining these data.

Controlled Analysis and Reporting: The data quality manager and stewards must govern data gathering, analysis and reporting. The IT department implements business rules and builds data structures to aid analyses. Multiple departments must share the ownership of reports whose data crosses departmental boundaries. A data savvy public information officer who is responsible for identifying and providing data to the public is highly desirable.

Data must be housed in a central data warehouse. There must be clear and available documentation of the enterprise data model and all related processes. Continually maintained metadata includes a data dictionary, business rules, report definitions, calendars, database schemas, and source code.

Although analysis and reporting are at the end of the data pipeline, incorrect processes and decisions at this stage bring serious consequences. Not a day passes without some report of conflicting statistics or inaccurate statistics resulting in funding mistakes, mislabeling and wasted resources.

Security and Confidentiality: Data security requires infrastructure components like authentication, access controls, firewalls, backups, antivirus, anti-spyware and intrusion detection. The agency must have a security officer who coordinates the agency's security plan.

All data workers must watch for possible breaches of confidentiality. More than just watching

for the disclosure of names and IDs, the agency must control the reporting of data representing small numbers of cases so an individual's identity cannot be inferred or traced.

Obviously bad data security can result in stolen identity that can open the education establishment to damaging lawsuits; but there are many more insidious consequences. Consider the accidental deletion of a small group of student assessment records in a data warehouse file by a data steward. Such an accident cannot only skew a report it can cause an inappropriate intervention to be scheduled for a class or an entire school.

Resource Management: Governance of data resources includes hardware, software, institutional knowledge and personnel. The data governance board, IT staff, data quality management, and stewards all must contribute to data plans and budgets.

Competing objectives, changing requirements and new technologies must be managed. The data governance board needs to collaborate with others to decide on how to adapt to unforeseen needs or unanticipated complications. Change management processes and life cycles for planned and ad hoc projects must also be defined and governed.

Infrastructure may be at the bottom of the data food chain but if it's inadequate the results can range from system crashes during online testing sessions to a failure to meet important deadlines like yearend report cards or accountability reports. If infrastructure is over allocated one can see a significant waste of expensive resources.

Data Use and Accessibility: Beyond security and confidentiality requirements, an essential element of providing quality data is a data governance plan that exacts a manageable balance between data usage/access requirements and timely and secure data distribution processes. Ensuring that data consumers have a consistent understanding of the definition and purpose of each accessible data element extends the utility and value of quality data. That understanding defines the potential value and limitations of available data while providing a higher probability of consistent, accurate reporting and analysis.

For More Information Contact:
Maureen Matthews
202-336-7023
maureenm@ccsso.org
www.ccsso.org/eimac