

# A Comprehensive pK-12 Data Model

October 2005

## Rationale

In addition to the evolving school, district, and state accountability and data-driven decision-making systems, the passage of the *No Child Left Behind Act* (NCLB) increased the demand for more immediate access to high-quality education data at all levels of pK-12 education enterprise. The bulk of information needed to respond to these information needs and other federal reporting requirements is initially captured in local schools and district offices to support local operations and decision-making. A subset of this data is then moved through state information systems to the United State Department of Education (USED). The dependence of federal, state, and local decision makers on the quality and availability of education data has never been higher.

A comprehensive data model, made available as soon as possible, would provide substantial support for current state and local data system developments. The development of such a model should include a thorough evaluation of data needs at the local, regional, state, and federal levels. Many districts and states want to design data warehousing and data collection sites with the purpose of greater tracking of cross-sectional and longitudinal data and many of them face challenges in designing a comprehensive data model, and others don't have sufficient resources to do so. Making a comprehensive pK-12 data model available to the education community, would save some states and districts time and resources, allow those with limited resources have access to a data model, and by providing a comprehensive menu of data options, would facilitate discussion at all administrative levels as to what should be included or excluded in a particular data system to be revamped or developed. To fulfill these goals, the comprehensive pK-12 data model must address data needs of key local stakeholders, including teachers, schools, and districts.

A complete picture of the pK-12 environment would not only facilitate the visualization of data transfer between applications within schools, but also conceptualize how to streamline reporting from the school to district, to regional, state, and ultimately, in a narrower content, to the federal level. In addition, this data model would facilitate the development of applications designed to provide administrators and educators with the tools they need to collect and analyze student learning and achievement, as well as other factors affecting them.

## Summary benefits of a comprehensive pK-12 data model:

- A full model WILL provide the core data structures to allow systems to be made operational.
- A full model WILL provide the core data structures needed for the shared federal data reporting burden to be addressed.
- A full model WILL provide a common vocabulary and framework for entities to have shared conversations about curriculum, data exchange, and standards.
- A full model WILL provide guidance on how data on classroom activities and programs relate to other components of the data system. This is an area of critical need in promoting effective decision-making regarding student learning and achievement.

## Identifying the Target Audience for a Comprehensive Education Data Model

An expert panel convened by SIFA agreed that in addition to states and districts the target audience would include developers and data architects that engage in building education data models. The most difficult

task in creating a comprehensive education data model is having a starting point that represents the best thinking to date. The categorization of the data sets and the comparison to the existing examples will give the architects and developers a powerful starting point to build out the section models that the states, LEAs, and schools can utilize.

The target audiences for the comprehensive education data model include:

1. Technologists in the state and local education organizations that are designing and planning their data warehouses, data analysis, and decision support systems. This model will allow them to start from a baseline covering 80% of the needed areas. These technologists can then focus their attention on customization and details particular to their educational environment. This will save them time and money.
2. Business managers and administrators who are writing and evaluating Requests for Proposals (RFPs). Having a sample model laid out for them as an example to work from will help them to create more effective RFPs. It will also allow them to evaluate vendor responses more rigorously and, in the end, get products that are closer to their needs. Although this is ranked as the third target audience, it may be the audience that has the biggest resource impact on education organizations in terms of saving time and money and moving states, LEAs, and schools forward quickly.

Vendors, while not a primary target audience, will be impacted in that they help education organizations build out their data systems. The modules and models that will be included in the comprehensive education data model will allow the vendors to ensure their products serve the core needs of their clients so that the education organizations do not need to backtrack and do more work to cover basic requirements.

### **Identifying Essential Data Model Components**

The expert panel convened by SIFA identified the essential components to be included in the comprehensive education data model that should be developed in a proposed follow-on project. These include:

- *Data Dictionary* – All of the essential data elements should be identified and defined.
- *Entity-relationship diagram (ERD)* – A graphical representation of the essential data elements and entities and their relationships.
- *Security/sensitivity documentation* (FERPA, Policy, law, minimum n-value) – A discussion of the security and privacy requirements relating to the essential components of the data model.
- *Time representation* – A discussion of how granular data elements at one level relates to aggregate data elements at higher levels, as well as issues and/or changes needed at all levels.

### **Preliminary Data Model Sections**

As a result of the review of the sample data models and discussion by the expert panel, preliminary data model sections were identified which were greatly influenced by the NCES Handbook organization. This was done to organize the information in a linear way and is considered to represent “functional groups of information” (i.e. HR, finance, etc. information).

Based on the NCES Data Handbook, there are four “organization” domains (SEA, IEU, LEA, School) and two “person” domains (Student, Staff). In this model, each of these domains would become core tables or groupings. There was some discussion about the use of Program and Class as “organization” domains, but the panel determined that that task should be addressed as part of the follow-on project. While these

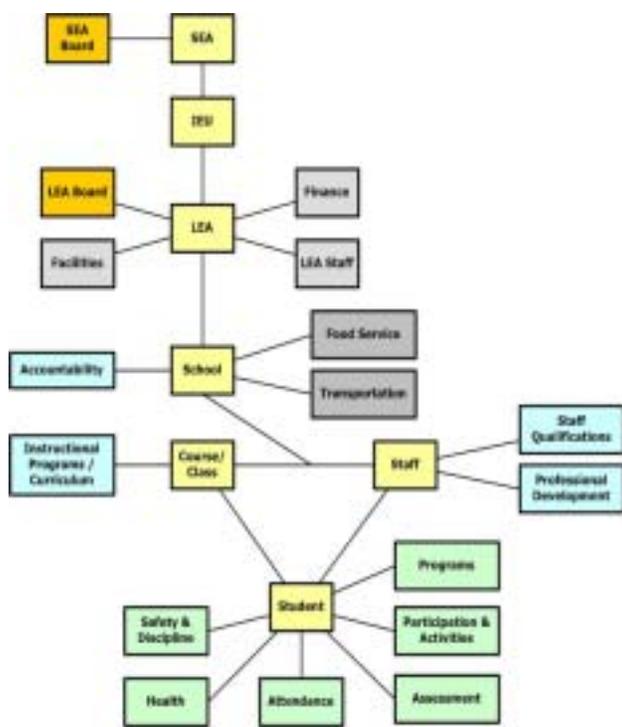
elements function much as a school functions in a data model, they are complex entities and may be best treated as functional domains.

### Proposed General Framework Diagram

In response to the complexity of existing data models at federal, SEA, and LEA levels, the expert panel generated a simplified diagram to represent a more accessible view of education data model categories and sections. In this diagram: yellow represents the major “ownership” institutions, gold legislative or policy oversight entity, grey various operational data areas, aqua staff and instructional components, and green student specific data components.

This does not represent the only “correct” way to structure a state, LEA, or school data system, but is meant as a framework informed by existing models, in which the most common relationships and ties between data sets can be represented. As each agency would be adopting sections of the framework, it would need to make the definitive relationship decisions that would represent its unique data needs and utilization requirements.

This modularization and detailing for each of the sections would constitute the primary output of the next phase of this project, and be the primary value-add to states, LEAs, and schools struggling with the challenging prospect of creating a thorough and fully enabled data system.



### Suggested Next Steps

With the lack of a comprehensive education data model available to states and districts today, the logical next step is for a large-scale stakeholder engagement strategy around the development of just such a model for federal, SEA, and LEA utilization. This future deliverable would not only facilitate local decision-making, but add to overall data quality by aligning data to be aggregated at the source. It would produce the “cleanest” data available. It is critical that this engagement would involve all relevant stakeholders in data utilization at the national, state, and local levels, including school and classroom levels. This includes educators, policy makers, and researchers, as well as consultants and vendors.