

## **Recommendation Paper**

### *Early Learning Challenge Data System: Data Standards*

*Race to the Top Project Technical Team:  
NM Children, Youth, and Families Department  
NM Department of Health  
NM Public Education Department*

## **Data Standards for the Early Learning Statewide Longitudinal Database System**

*7/2/2014*

## Table of Contents

Table of Contents.....	2
Executive Summary.....	3
<i>Option 1 – Common Education Data Standards (CEDS).....</i>	<i>3</i>
<i>Option 2 – Creation of a New Mexico data standard.....</i>	<i>3</i>
<i>Recommendation: Option 1 – Common Education Data Standards (CEDS) version 4.0.....</i>	<i>3</i>
Full Report – Detailed Information.....	4
<i>Option 1 – Common Education Data Standards (CEDS) version 4.0.....</i>	<i>4</i>
<i>Pros.....</i>	<i>6</i>
<i>Cons.....</i>	<i>6</i>
<i>Option 2 – Creation of a New Mexico data standard.....</i>	<i>6</i>
<i>Pros.....</i>	<i>6</i>
<i>Cons.....</i>	<i>6</i>
<i>Recommended Approach: Option 1 – Common Education Data Standards (CEDS) version 4.0.....</i>	<i>7</i>

## **Executive Summary**

The Race to the Top Technical Team was established to examine possible options for design for the Early Learning Challenge Longitudinal Data System and make recommendations for approval by the Data Governance Committee. The agencies have previously met on several occasions to discuss topics including possible data standards and on July 2, 2014; the technical team convened and decided on a preferred option for their data standard.

### **Option 1 – Common Education Data Standards (CEDS)**

Under this option, the agencies will agree to share data in a way that aligns it to the Common Education Data Standards (CEDS). CEDS is a national standard that was designed to facilitate the exchange, comparison, and understanding of data across systems, agencies, and states. CEDS provides a common vocabulary and structure for data including definitions, naming, options lists, data types, and other technical specifications.

### **Option 2 – Creation of a New Mexico data standard**

Under this option, a completely new data standard would be created by the agencies. The table and column names for each data item would have to be created and agreed upon by each agency. The group would have to decide upon the data types and the option list for every item as needed. They would also have to create and agree upon the methodology for structuring the tables and linking the data.

### **Recommendation: Option 1 – Common Education Data Standards (CEDS) version 4.0**

The unanimous consensus of the technical team is that the best option is to use the Common Education Data Standards (CEDS) as the foundation for our data system.

## Full Report – Detailed Information

### Need

The three agencies involved in creating a shared data system need to establish the rules by which they will align their data so that meaningful interpretations of the shared data can be generated. There are different approaches that can be taken in the creation of these rules and the continuation of development of the system is contingent upon agreement on the methodology so that the alignment work can begin.

### Summary

The Race to the Top Technical team is a collection of representatives including business analysts, database administrators, and application developers. These representatives from each of the involved agencies have the experience to be able to understand and represent their agency's capabilities and needs, listen to the other agencies, determine how to combine all of those capabilities and needs in a way that will effectively serve the project's needs, and present a recommendation.

Rules that are needed to align our data systems include:

- Naming conventions - Allow systems to agree to all call an item DateOfBirth rather than one agency providing DOB and another providing BirthDate.
- Options lists - Give the valid names for items to ensure that the data in each item is sent consistently and rather than having Male, M, and 1 - Male as values in a gender column.
- Definitions – Provide shared understanding that Grade Level represents the child's level in school such as Kindergarten and that Grade Earned represents the letter grade associated with a particular class such as a B+.
- Data types – Defined data types for each item ensure that characters are not unintentionally cut off the end of a value because it is too long and leading zeroes are not unintentionally removed from an identifier being changed from text into a number.

The agencies have all had the opportunity to discuss their ideas internally and had met between agencies on different occasions to discuss possibilities. On July 2, 2014 the technical team met and made a formal decision about their recommended approach.

### Option 1 – Common Education Data Standards (CEDS) version 4.0

Under this option, the agencies will agree to share data in a way that aligns the data to the Common Education Data Standards (CEDS). CEDS is a national standard that was designed to facilitate the exchange, comparison, and understanding of data across systems, agencies, and states. The CEDS website can be found at <https://ceds.ed.gov/Default.aspx>. CEDS provides a common vocabulary and structure for data including definitions, naming, options lists, data types, and other technical specifications. The CEDS collaborative also produces sets of use cases called "Connections" that list the data elements that would be useful in creating reports about specific topics. An abbreviated example of the CEDS element representing Sex is provided in Figure 1.

The CEDS collaborative also provides a sample normalized physical database structure that systems can use to implement CEDS. They have provided SQL scripts that can be used to create the database tables and populate many of the tables that hold the option sets.

**Sex**

**Definition**  
The concept describing the biological traits that distinguish the males and females of a species.

**Usage Notes**  
Postsecondary: As defined in IPEDS

**Option Set**

Description	Code
Male	Male
Female	Female
Not selected	NotSelected

**Related Entities and Categories**  
 Early Learning -> EL Child -> Demographic  
 Early Learning -> EL Staff -> Demographic  
 K12 -> K12 Staff -> Demographic  
 K12 -> K12 Student -> Demographic  
 K12 -> Parent/Guardian -> Demographic **NEW!**

**Related Connections**  
 Early Learning -> Program Entry  
 Early Learning -> Staff Quality  
 Early Learning -> Workforce Development  
 K-12 -> EDFacts  
 K-12 -> High School Generated Transcript  
 K-12 -> LEA-to-LEA Student Record Exchange  
 K-12 -> LEA-to-SEA Student Record Exchange  
 K-12 -> Teacher Compensation Survey

**CEDS Element ID**  
000255

**Element Technical Name**  
Sex

**XML**

```
<xs:simpletype name="Sex">
  <xs:annotation>
    <xs:documentation>Usage: Sex</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:token">
    <xs:enumeration value="Male">
      <xs:enumeration value="Female">
        <xs:enumeration value="NotSelected"></xs:enumeration>
      </xs:enumeration>
    </xs:enumeration>
  </xs:restriction>
</xs:simpletype>
```

**URL**  
<https://ceds.ed.gov/CEDSElementDetails.aspx?TermId=5255> ([Email this link](#))

Figure 1

Deviations from the CEDS standard will be proposed when the agencies agree that there is a defined need for additional data that is not included in CEDS or when the agencies collect data in a manner that is substantially different from CEDS. The CEDS physical model is designed in a way that allows flexibility for storing elements that would be required for our Race to the Top purposes that might not normally be found in an educational data model. For example, the tables for Person and Organization could accommodate a variety of providers. So even in cases where deviation from the entity/element concept is required, the underlying physical model may not require changes. Any needed modifications will be documented and reflected in the New Mexico documentation of our data system design.

During the team's ongoing analysis, it is expected that we will uncover situations where CEDS has a structure in place to hold data that New Mexico does not collect. We also expect to find data items that the agencies feel are not relevant or not prudent to share based upon reporting needs and privacy requirements. After we complete our initial analysis, the team will be better situated to understand the extent of the fields that will not be populated and will make a recommendation at that time to either remove those items from our data dictionary or to include them in the model and not populate them.

Additionally, the agencies feel that it is important to not create a situation where the requirements are constantly shifting, especially during the initial development that is occurring under the Race to the Top grant. For this reason, we are recommending a baseline at the current version of CEDS, 4.0, as the standard. If the team determines that there is a compelling reason to adopt specific changes from a future release, we will consider requesting the revision and propose a reasonable timeline for implementation of the change. It is expected that most of these changes would occur after the data system is in production, in response to specific future business requirements.

#### **Pros**

- Well established standard that has been vetted by a national cross functional group
- Documentation of the baseline has already been created and would just need modifications based on New Mexico's implementation
- Can leverage the existing use cases as a starting point for several future reports
- Immediate availability of detailed element level definitions
- Sample physical model can be created in minutes utilizing existing scripts
- CYFD has been designing EPICS with intent to be CEDS compliant
- STARS, the main longitudinal database for PED, is CEDS compliant

#### **Cons**

- We may have data that does not align well with the model
- The database model is extensive and we may not be utilizing all the available fields
- New versions are released regularly, effort would be required to maintain our system at the newest version

The information including pros and cons has been discussed by the group. The agreed upon baseline at version 4.0 combined with our commitment to controlled modifications of the structure as required to meet New Mexico's needs would mitigate most of the identified cons.

### **Option 2 – Creation of a New Mexico data standard**

Under this option, a completely new data standard would be created by the agencies. The table and column names for each data item would have to be created and agreed upon by each agency. The group would have to decide upon the data types and the option list for every item as needed. They would also have to agree upon the methodology for structuring the tables and linking the data, including which items would support one to many relationships.

#### **Pros**

- Custom to New Mexico, we can design precisely what we want to have and how we want it structured
- Minimalistic implementation could include only the fields we want to share and keep the overall structure simple

#### **Cons**

- Requires extensive collaboration during the creation of the data dictionary
- Time intensive to create and document every element

## **Recommended Approach: Option 1 – Common Education Data Standards (CEDS) version 4.0**

The unanimous consensus of the technical team is that the best option is to use the Common Education Data Standards (CEDS) version 4.0 as the foundation for our data system. By adopting this already complete, nationally recognized data standard; the team can begin focusing on the analysis of their existing systems against this standard and understanding the mapping or transformations that will be required to align their data. It will allow us to leverage work being performed across the country such as the Connections use cases and facilitate possible future collaboration with other CEDS compliant systems, agencies, or states.