

## **Recommendation Paper**

### *Early Learning Challenge Data System: Person Unique ID System*

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

## **Person Unique ID System for the Early Learning Statewide Longitudinal Database System**

*9/3/2014*

## Table of Contents

Table of Contents.....	2
Executive Summary.....	3
<i>Option 1 – Create a custom system to issue and manage person identifiers.....</i>	<i>3</i>
<i>Option 2 – Utilize a third party vendor to issue statewide identifiers.....</i>	<i>3</i>
<i>Option 3 – Utilize an existing deterministic identifier or combination of identifiers.....</i>	<i>3</i>
<i>Recommendation: Option 1 – Create a custom system to issue and manage person identifiers.....</i>	<i>3</i>
Full Report – Detailed Information.....	4
<i>Option 1 – Create a custom system to issue and manage person identifiers.....</i>	<i>4</i>
<i>Pros.....</i>	<i>5</i>
<i>Cons.....</i>	<i>5</i>
<i>Option 2 – Utilize a third party vendor to issue statewide identifiers.....</i>	<i>5</i>
<i>Pros.....</i>	<i>5</i>
<i>Cons.....</i>	<i>5</i>
<i>Option 3 – Utilize an existing deterministic identifier or combination of identifiers.....</i>	<i>6</i>
<i>Pros.....</i>	<i>6</i>
<i>Cons.....</i>	<i>6</i>
<i>Recommended Approach: Option 1 – Create a custom system to issue and manage person identifiers.....</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 including July 9<sup>th</sup>, July 16<sup>th</sup>, July 30<sup>th</sup>, and August 13<sup>th</sup> to discuss our recommendations for a Unique Identifier system for people in New Mexico systems.

### **Option 1 – Create a custom system to issue and manage person identifiers**

Under this option, the agencies will design and PED will develop a new system to issue, store, and manage a new identifier solely for the purposes of matching person data across agencies for the Race to the Top database. The system will need to have the capability to recognize existing people, issue ids for new people, allow for splitting and merging of ids, and allow for the resolution of near matches.

### **Option 2 – Utilize a third party vendor to issue statewide identifiers**

Under this option, a third party vendor that currently is currently utilized for the issuance of identifiers for some programs could be contracted with to provide an identifier for all participating agencies. The existing system has the capability to recognize existing people, issues ids for new people, allows for splitting and merging of ids, and allows for the resolution of near matches.

### **Option 3 – Utilize an existing deterministic identifier or combination of identifiers**

Under this option, an id or combination of ids that already exist in the systems could be relied upon to provide the matching. Examples of ids that could be used include social security numbers, driver's licenses numbers, and state issued student ids.

### **Recommendation: Option 1 – Create a custom system to issue and manage person identifiers**

The unanimous consensus of the technical team is to create a new system to issue identifiers specifically created for our matching purposes utilizing criteria and algorithms that work with the agencies' existing data systems and data.

## Full Report – Detailed Information

### Need

The three agencies involved in creating the shared data system need a means of merging the information about people (primarily children and staff) in their individual systems with information about the same individuals from the other systems. It is important that data elements from the different agencies can be linked with a high level of confidence that all the data are for a single individual. It is also important that the individuals are de-duplicated, that is that an individual that appears in multiple systems will only have one combined record when viewing from the merged system. There are different approaches that can be taken in the creation of these ids and the development of the unique id system is contingent upon agreement on the approach to take.

### 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.

Requirements that are needed surrounding person unique ids include:

- Inquiry – Determining if an individual already exists within the unique identifier system and what that individual's identifier is.
- Resolution – The ability to return the information that a person might already exist in the system and provide a mechanism for the resolution of whether or not they do exist and which person they are if there are multiple possibilities.
- Creation – The issuance of a new, not previously issued identifier when it is determined an individual is not currently in the system.
- Splitting – The ability to issue new identifiers when two individuals have been inadvertently issued a single identifier as well as communication back to the agencies about the new identifiers.
- Merging – The capability of merging two identifiers into one identifier when it is determined that an individual was issued two identifiers and the communication back to the agencies about the correct identifier.
- Security – Ensuring that only authorized applications/people have the ability to access the information inside the identifier system.

The agencies have all had the opportunity to discuss their ideas internally and had met between agencies on different occasions to discuss possibilities. On September 3, 2014 the technical team met and made a formal decision about their recommended approach.

### Option 1 – Create a custom system to issue and manage person identifiers

Under this option, the agencies will design and PED will develop a new system to issue, store, and manage a new identifier solely for the purposes of matching person data across agencies for the Race to the Top database. The system will need to have the capability to recognize existing people, issue ids for new people, allow for splitting and merging of ids, and allow for the resolution of near matches.

During the team's ongoing analysis, it has been discovered that there are certain known data quality issues relating to the data that will be used to match people. For example, some systems keep only a middle initial while others a full middle name. Additionally, some systems append the generational suffix to the last name while others store it separately. We can address these issues in a custom system by implementing a name standardization routine.

### **Pros**

- We can control the releases of new versions and allow time for agencies to adjust their interfaces if needed.
- We can completely customize our matching algorithm to give weight to the most reliable and commonly collected data and utilize questionable or scarce data only in secondary matching routines.
- We can design the interface mechanisms that work best for our agencies (flat files or web, batch or single record).
- We can issue an id that is not utilized for any other purposes and therefore would not need to be displayed in the agency systems. This adds a layer of security since data only identified by this identifier could not be tied directly to other data sets or an individual.
- Agencies can implement without modifications to their existing data systems or databases by utilizing a crosswalk approach.
- Once developed, the system can issue identifiers for additional programs and increased numbers of participants for no additional cost.

### **Cons**

- The matching algorithms and other coding required to create a unique identifier system are complex and will require a significant investment of development time.
- Agencies will not be able to take advantage of existing interfaces and system knowledge.
- The creation of a new system will involve an up front development cost, currently estimated at approximately 1000 hours of process, database, and system development.

The information including pros and cons has been discussed by the group. We could consider utilizing a contractor to create our system if resources are over-allocated. We can also make efforts to make the interfaces with the system similar to existing systems to minimize the learning and development curve for agencies that already have existing identifier systems.

## **Option 2 – Utilize a third party vendor to issue statewide identifiers**

Under this option, a third party vendor that currently is currently utilized for the issuance of identifiers for some programs could be contracted with to provide an identifier for all participating agencies. The existing system has the capability to recognize existing people, issues ids for new people, allows for splitting and merging of ids, and allows for the resolution of near matches.

### **Pros**

- The programs already using the vendor could potentially use the already issued identifiers and existing interfaces
- The matching algorithms and other coding required to create a unique identifier system are complex so utilizing an existing system could save time.

### **Cons**

- There are ongoing costs (annual and/or per id issued) associated with utilizing an external vendor that were not budgeted for in our initial application and raise questions about sustainability after the grant conclusion.
- We propose to use a single system for issuing identifiers for staff and children, the existing identifier system is not utilized in this way.
- Utilizing a third party system limits the ability of New Mexico to customize the matching process based on known data issues.
- Utilizing an existing system limits our ability to customize the security roles and the interface for retrieving the identifiers.

### **Option 3 – Utilize an existing deterministic identifier or combination of identifiers**

Under this option, an id or combination of ids that already exist in the systems could be relied upon to provide the matching. Examples of ids that could be used include social security numbers, driver's licenses numbers, Medicaid numbers, licensure numbers, and state issued student ids.

#### **Pros**

- No new identifiers or systems to support the identifiers would need to be created

#### **Cons**

- Analysis demonstrated that we do not have any identifiers that exist in every system so combinations would be required.
- It is possible that a child could exist in multiple systems that do not share any identifiers and consequently they would be duplicated and their data unmatched.
- Asking agencies to collect additional identifiers to alleviate the above concerns would be burdensome to our service providers and require system modifications to our existing systems.
- This use of certain identifiers for cross agency data matching is controversial and/or statutorily prohibited due to security and privacy concerns.

## **Recommended Approach: Option 1 – Create a custom system to issue and manage person identifiers**

The unanimous consensus of the technical team is that the best option is to issue our own identifiers using a system that we create for the Race to the Top data matching purposes. By creating our own system, we can completely customize it to fit New Mexico's unique needs and design it to function with our existing systems rather than changing our systems to fit the needs of the unique id system. We can ensure a high level of accuracy in our matching and enhance the system to continually improve our matching as more information is gathered. We also add a layer of security by not storing or transmitting the program data connected to identifiers that are known to entities outside of the collaborative. Finally, the source code and system could be made available to other state agencies and initiatives that might have similar needs.