



---

## The Role of the Unique Student Identifier in Longitudinal Data Systems

---

**“Educators and policymakers recognize the value of better information as an essential tool for improving schools and longitudinal analysis is a key to success.”**

### Creating a Longitudinal Data System

When states collect highly relevant data and are able to match individual student records *over time*, they can answer the questions that are at the core of educational effectiveness. Longitudinal data, data gathered on the same student from year to year, makes it possible to follow students’ academic progress as they move from grade to grade.

Policymakers and educators need longitudinal data systems capable of providing timely, valid and relevant data. Access to these data gives teachers the information they need to tailor instruction to help each student improve, gives administrators the resources and information to effectively and efficiently manage programs, and enables policymakers to evaluate which policy initiatives show the best evidence of increasing student achievement.

**“Although each state’s education system demands may differ, the statewide unique student identifier must be an essential element in a longitudinal data system.”**

### Unique Statewide Student Identifier - An Essential Element

A unique statewide student identifier is a single, non-duplicated number that is assigned to, and remains with, a student throughout his or her educational tenure. A student who leaves the state and returns should be assigned his or her original number. A student identifier will allow the state to follow the progress of each student over time, from pre-kindergarten across campuses or districts within the state while ensuring his or her privacy. It also makes it possible to identify information about a single student across various data sources (e.g., enrollment, program participation, demographics and assessment) to evaluate the relationship between program participation and performance and to study student mobility patterns and evaluate the effect of mobility on performance.

School Districts and State Education Agencies understand that an accurate process of ensuring unique student record identification is the key element in a longitudinal data system, but have often struggled to achieve and maintain unique identities of students and respect student confidentiality. Manual creation of duplicate records, inability to process students in real-time and inefficient matching technology has plagued implementation of successful systems.

In building a statewide data infrastructure, the system should ensure that student records can be easily transferred, student privacy is protected, data definitions and requirements are clear to concerned, and the data system is organized in ways that facilitate data use and user-friendly reporting. The CPSI xDUID Unique Identifier System meets this need and delivers a highly sophisticated system to manage student identification data, conduct accurate probabilistic matching, and manage deduplicate student records.



“ The xDUID technology provides the needed foundation for longitudinal data systems based on the No Child Left Behind (NCLB) requirements. ”

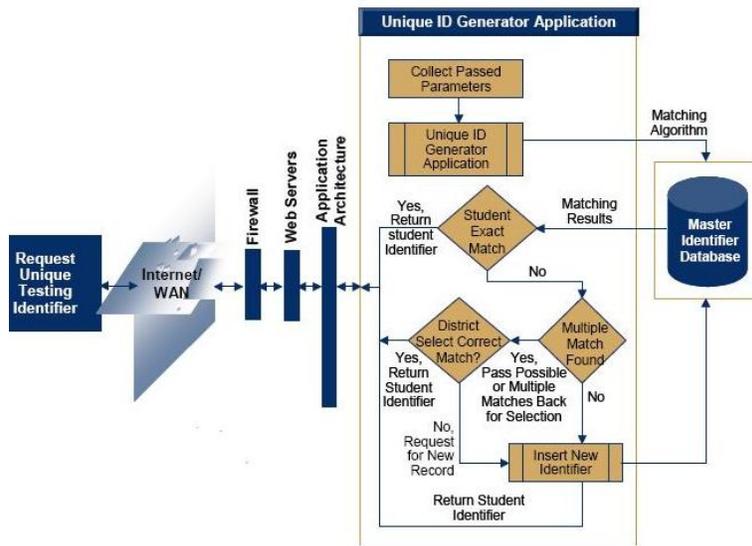
“The system should be compliant with the Schools Interoperability Frameworks (SIF) and Deliver on the Mandate for Vertical Reporting in K-12 Education”

### Schools Interoperability Standard (SIF) – A Backbone

SIF is recognized an important methodology in addressing the challenges posed by disparate IT systems and is a critical step toward improving access to necessary information. SIF is an industry-supported open standard that allows schools to choose compliant applications for managing a wide variety of data.

The xDUID is designed for seamless integration into state education administrative systems and it is fully SIF compliant supporting the SIF Student Locator Object. It will create and manage unique student identifiers (UID's) and a master identifier database of all students who have received public education services. The identifiers are unique and unchanging, enable compliance with state and federal regulations, and provide the link necessary to facilitate the use of a wide variety of data sources. As each unique identifier is created, a set of carefully selected student demographic attributes is stored with the identifier. The same identifier is retrieved as students change schools, change districts, or exit and re-enter the education system.

### The core of the application is the CPSI xDUID Match Engine



The CPSI xDUID Match Engine is used to evaluate selected attributes to determine if a student already exists in the system. In cases where the CPSI Match Engine cannot discern whether two or more records describe the same student, users are presented with the information needed to make a correct determination. If no existing record describes the student in question, a new UID is created and assigned. The algorithms contained in the Match Engine have proven successful in assigning over 1.9 million student identifiers.



In the xDUID, identifiers are unique and unchanging and are never deleted or reused. As each UID is created, a set of student demographic attributes is stored with the identifier. Attributes can be customized in the system.

**“The system should have robust and proven matching methods, accommodate multiple outcomes, and the system should allow an administrator to tune system threshold ranges that determine whether records are considered a match or need further review.”**

The matching technology contained in the CPSI Match Engine uses a two-pass system to quickly eliminate dissimilar records from consideration. The few remaining records are examined in detail for similarity. In some cases, a character-by-character review of each field is initiated.

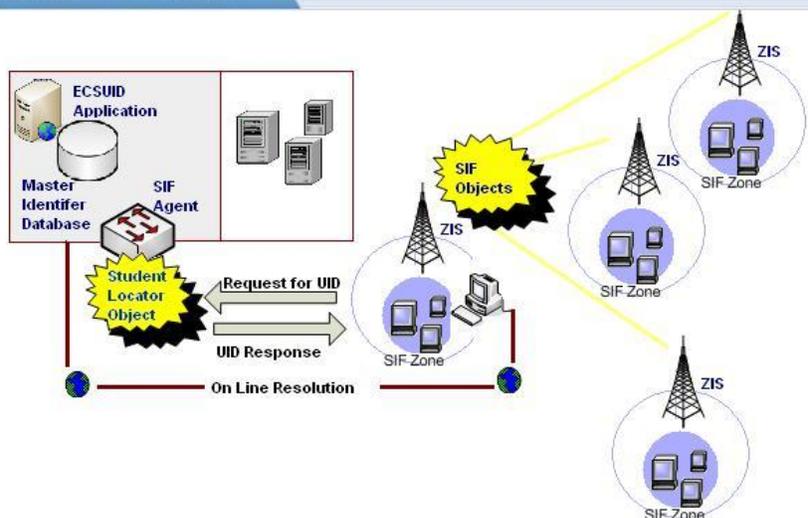
The CPSI Match Engine determines a normalized match score and applies the result to thresholds set for confidence limits among matches. This approach allows the system to determine their levels of for the special characteristics of text, numeric, and date elements.

The xDUID application is configurable with regard to the type and nature of attributes stored and used to describe students. Configurable weighting factors are used in calculations.

UID's can be assigned using either of three methods:

- Online using a web based user interface for an individual student.
- Using a batch upload method.
- Interoperability with other SIF compliant applications

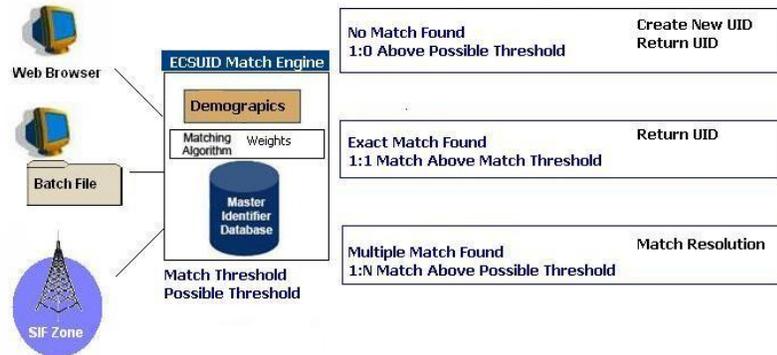
#### ECSUID SIF MODEL



In the SIF model, the Student Locator is the central object that supports the use of the unique identifier. The xDUID supports the Student Locator Object and acts as the provider of the object and data within the system. The xDUID resides at the state and utilizes the state Zone Integration Server (ZIS). The ZIS routes all requests for a UID to the xDUID and the xDUID SIF agent handles the request and submits the data to the xDUID Match Engine. The outcome of the Match Engine calculation is published back to the requestor.



Without regard to the mode of submission, the xDUID Match Engine will return one of three possible results:



1. Where the input data does not match any student, the application will return a no match condition and create and assign a new UID.
2. Where the input data matches a single existing student, the application will return the UID and details of that student.
3. Where the input data matches multiple students, or is similar enough that it may match one or more records, a list of the matching students will be returned.

### Match Attributes

Each student in the system is defined by a series of attributes. The xDUID can accommodate up to 19 attribute fields in the aggregate. Fields can be of type string, number or date. Attributes are both configurable in their nature as well as the weights applied in matching calculations. The 19 attribute fields cannot be deleted in the xDUID system; however, they can be marked inactive and, as a result, are not used or displayed in the application (effectively “turning them off”).

Attributes in the xDUID are provided in two categories, standard and custom. Standard attributes are part of the xDUID framework and are provided by default. Standard attributes are configurable for weight and use in the system. Up to ten user defined custom attributes are provided and are user configurable in the system.

Each of the standard and custom attributes can be one of three (3) types in the xDUID. The attribute type describes the use of the attribute within the system and its relationship to the xDUID Match Engine.



### Attribute Types

**KEY** – Two (2) attribute fields are required for the function of the core match engine. One field must be a string and the other must be a date. By default, the key string attribute is the student last name and the key date attribute is the student date of birth. Key fields must carry an assigned weight. Key attributes are displayed on screens and reports for match resolution. Key fields are immutable and cannot be changed for the life of the application. Key attributes are always included in the CPSI Match engine calculations.

**“The system should be flexible enough to permit changes in and additions to attributes and other information fields as needed to improve efficiency.”**

**Comparison** – These fields are included in the match engine calculations and must carry an assigned weight. Comparison fields are selected as identifiers, that is, they change very little. Comparison attributes are displayed on screens and reports for match resolution. Comparison attribute fields are user configurable relative to data entry, use, and weights.

**Helper** – These fields are not included in the match engine calculations and are not assigned a weight. Helper field are not used as identifiers because they are known to change. Helper attributes are displayed on screens and reports as an aid in match resolution but do not contribute to match scores. An example of a helper field commonly used is student grade.

The Administrator can configure aspects of the attribute fields through XML configuration files and web UI screens. The user interface will allow the selection of weights, the type of attribute fields, and match thresholds for custom attributes. Certain aspects of default attributes and the nature of the default attributes can also be configured by the Administrator. User configuration will not modify the number of supported attribute fields, but will allow for assignment of comparison fields, adjust weights, and provide display/use information.

### **Security**

The xDUID application restricts access to identifiers and associated attributes on the basis of user role and user organization to prohibit disclosure of personally identifiable information concerning a student in the master identifier database. *Users can access data for individual students only within their sphere of responsibility.* A limited number of users can access data for all students, while most users can access data only for a particular district, school, or set of schools within a district.

**“Security is an important consideration in the system and should be role based.”**



**“A key element in the successful implementation of a unique identification system is the process of preventing and managing duplicates.”**

### **De-duplication**

Duplicates may arise, without regard to the accuracy of the CPSI Match Engine, when errors are made in data entry or erroneous information is provided to the system. Periodic de-duplication of the xDUID master database is an important operational function. The process for a system wide de-duplication process is built into the xDUID framework and the system recognizes the dynamic nature of the unique identification process and provides a web user interface to manage and evaluate potential duplicates in the master identifier database.

**“The system should not delete a record and retain history of all attribute values, aliases, and modifications”**

### **Linking Students**

When duplicates are identified, the xDUID provides a linking mechanism to link one student to another to accommodate the duplicate while maintaining the integrity of historical records. Once a student is established with a UID, the student cannot be deleted nor the UID reassigned. An Administrator can link students that have different attribute information, but are known to be the same individual. Linking records establishes a method to define multiple records as the same student and preserving the existing UID's. Among linked records, only one record is designated as primary at any given time, however, all of the historical information is retained to support retrieval of information without the requiring districts to edit or alter data. The system will maintain an audit trail of linked students.

**“The system should have a flexible UID format and an integrated check digit to validate a properly formatted UID.”**

### **UID Format**

The format for the UID generated by the xDUID is configurable to any length between 7 and 12 digits. xDUID UID's are randomly generated and have the following characteristics:

- Contain random numbers – not sequential.
- Do not begin with a zero (0) or end with a zero (0).
- Do not contain a sequence of 3 or more repeating numbers.
- Use only numbers.

The UID contains a check digit constructed using Modulus 11 calculation in accord with National Serials Data Program methods.



**“The system must be state of the art technology, flexible to meet changing regulations, and scaleable for future needs”**

## Application Architecture

The application architecture provides many features that will enable states to respond to change rapidly. Additionally, with Microsoft .NET as a foundation, its *n*-tier structure allows for development and modification of individual components, including the introduction of new technologies (or the replacement of outmoded ones) in each tier, with minimal impact on components provided by other tiers.

The xDUID application framework makes best use of the following principles:

- *Open standards.* The ability to exploit current and emerging standards including XML. SIF interoperability as a base capability.
- *Component architecture.* The resolution of an application into an *n*-tier structure that allows for distributed use of component services and encourages reuse
- *On-demand scalability.* The ability to adjust the required scale on demand, without disruption of ongoing operations. The product platform also fully supports web farms that use either hardware or software load balancing and session management is fully supported. The result is the ability to scale to higher traffic levels without application modification.
- *Built-in security.* Exceptionally strong, role-based security protection with binding privacy policies

The solution architecture was developed in conformity with numerous best practices:

- The solution components mimic business functions. The component-based design allows distribution of functionality as needed.
- The solution is highly granular and loosely coupled, service-oriented and component-based, the design accommodates any degree of granularity required.
- The solution supports extensibility and scalability. The component-based design and inheritance support extensibility while clustering and load-balancing support provide scalability.
- Tools are based on application architecture and have been matched to the architecture for each tier.



**“Schools, districts, and most important, the children they educate, benefit much more from technology when all of their computers, programs, and networks share data simply and securely.”**

An important part of K-12 education is the ability to collect and analyze data from a variety of resources, including at the state level. An essential element of longitudinal analysis and vertical reporting is the use of an accurate SIF enabled unique student identifier. This technology helps meet the challenge of integrating data from districts, their schools, and the state into a meaningful timely and consistent manner.

The combination of the xDUID Match Engine as the application backbone, the robust attribute set, and a fully functional SIF compliant model provides the ideal platform to implement a unique student identifier system and support longitudinal analysis and vertical reporting now and in the future.

Technology trends and policy initiatives in K-12 education have helped fashion a new generation of data-driven stakeholders that operate in real-time. Students, educators, administrators, policy makers and parents, have heightened expectations for the ways in which student progress is tracked and impact is measured within our school systems. The xDUID technology plays an important role in enabling delivery on the promise of NCLB and meeting the 21<sup>st</sup> century data-driven challenges.