

Linking Records Across Data Systems, Part 1: The eScholar Student UID System

A How NCLDS Works Brief for NCLDS Contributors and Stakeholders

Government Data Analytics Center

Version 1.3a December 2023

Table of Contents

The <i>How NCLDS Works</i> Series	2
1. Purpose	3
2. eScholar Student UID System Overview	3
3. UID Assignment Methods	3
4. UID Assignment Overview	4
5. North Carolina Entities that Assign UIDs	7

The *How NCLDS Works* Series

This brief is part of a series that provides details for North Carolina Longitudinal Data Service (NCLDS) users, NCLDS Data Contributors, and other stakeholders about how various technical and procedural aspects of NCLDS and the systems that contribute data to NCLDS work. The briefs focus on aspects that are not easily explained in a paragraph or two.

Each brief has been written in a way that we hope will make it accessible even to audiences without data, analysis, or technical backgrounds, but please share feedback with us about how we can make the briefs more accessible. We are also open to suggestions for other topics you would like to see covered. We can be reached at NCLDShelp@nc.gov.

Currently Available Briefs

- Linking Data: eScholar Student UID
- Linking Data: eLink Entity Resolution
- Linking Data: Workforce Data

Planned Briefs

- Linking Data: Prospects for *Ad Hoc* Matching
- Using the Public Version of the NCLDS Data Dictionary
- Making Data Requests
- Reviewing and Approving Data Requests
- Fulfilling Data Requests
- Reviewing Products Created by External Partners with NCLDS Data
- Cross-Sector Governance of NCLDS
- Security and Privacy
- Creating Practitioner Portals
- NCLDS Cross-Sector Learning Goals

1. Purpose

One key component of the usefulness of NCLDS is the availability across NCLDS data sources of reliable and up-to-date **record-level¹ identifiers**. Identifiers help NCLDS connect separate pieces of data to each other (for example, a person's high school academic outcomes and postsecondary course enrollment). Without these identifiers, important data may not be included in analyses that assess the value and impact of policies, programs, and supports.

This brief highlights one of the identifiers critical to understanding education-related outcomes—the North Carolina Department of Public Instruction's (NCDPI) eScholar Student Unique Statewide Identifier (UID). The brief also lists the Contributors (entities that make data available to NCLDS) that attach UIDs to their data records.

2. eScholar Student UID System Overview

The NCDPI eScholar Student UID System is used to generate and assign a UID for every student who participates in the North Carolina public education system at any point in time (pre-Kindergarten through postsecondary). The UID enables reliable matching of student-level records over time and across data systems. The UID follows a student across time and location and remains valid even if the student moves out of state but later returns to North Carolina. The Student UID system is built on the eScholar UniQ-ID® product and is operated and managed by NCDPI.²

3. UID Assignment Methods

There are three methods that can be used to submit records to the eScholar Student UID System for assignment of a UID: **batch file upload** via a user interface, **automated file processing (AFP)** via secure file transfer protocol (SFTP) to the database application server, and **web service integration** (see **UID Assignment Terminology** box). Each of these methods either has been used in the past or is still currently being used by at least one NCLDS Contributor for UID assignment. Two of the three—AFP and webservice integration—are semi- to fully-automated methods for assigning IDs.

Record Submission Terminology

File Upload: Manual upload of a text file containing student records. A *batch* file upload contains multiple student records.

Automated File Processing: Submission to a state-hosted secure server location of a text file that contains one or more student records; the eScholar Student UID System retrieves the file from the secure location. The process can be automated to reduce human interaction with the data.

Web Services: A third-party software-enabled upload of a text file with a single student record.

¹ For NCLDS, a **record** typically refers to data linked to a specific individual (e.g., a student or worker)

² More information about the system is available at <https://www.dpi.nc.gov/data-reports/common-education-data-analysis-and-reporting-system-cedars/student-uid-system#Recordings-1030>

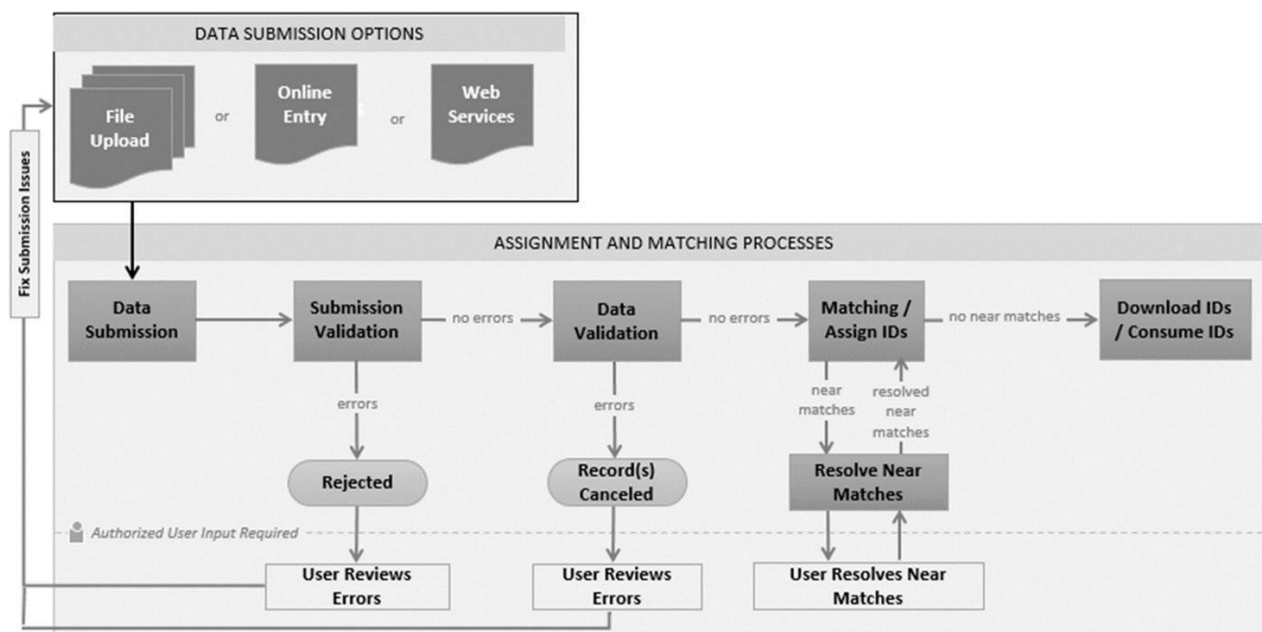
Records can be submitted to the system in one of two grouping formats. When using either the batch file upload or AFP method, the file sent can contain multiple records for UID assignment. When using the web service integration method, the files can contain only one record per file.

Based on information provided by current Contributors, their frequency of file submission can range from nearly constant to only twice annually.

4. UID Assignment Overview

The eScholar Uniq-ID® product uses an algorithm to match submitted student records to existing records and to then assign a random unique 10-digit numeric UID value to the record.³ To assign UIDs, the system goes through three phases of processing, including **submission validation**, **data validation**, and **matching and assignment**. Figure 4.1⁴ illustrates how submitted records move through the UID assignment process in the Student UID System (the process starts in the “Data Submission Options” box and ends in the “Download IDs / Consume IDs” box).

4.1 eScholar Uniq-ID Match and Assignment Data Flow



The **submission validation** phase checks the file of submitted records to determine if it meets the necessary specifications to be processed by the system. If the file passes the submission validation phase, the system then performs **data validation** for each submitted data element in the file, based on eScholar system and data owner-specific rules.

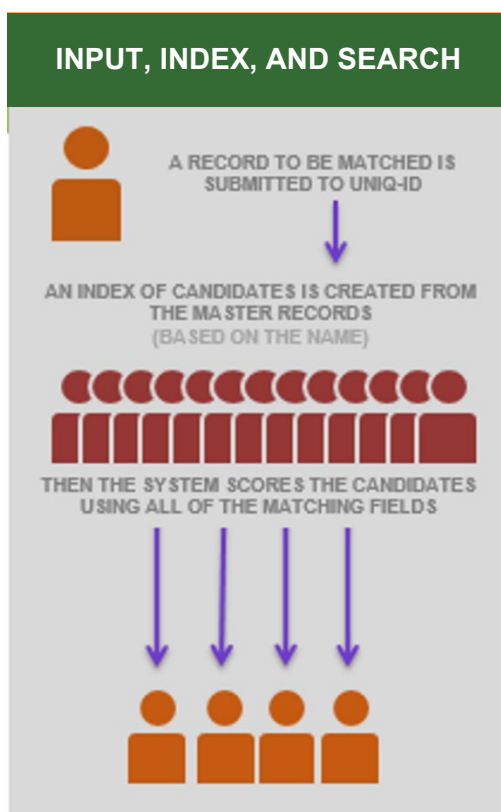
³ If a new record is clearly matched with existing records already in the system, the new record is assigned the UID already associated with those existing records.

⁴ Figures 4.1, 4.2, and 4.3 were developed by eScholar.

Once the submission validation and data validation steps are complete, the system sends each record contained in the file through the matching algorithm to **assign a match score** (details below) and **match the record**, if possible, to existing records, as follows:

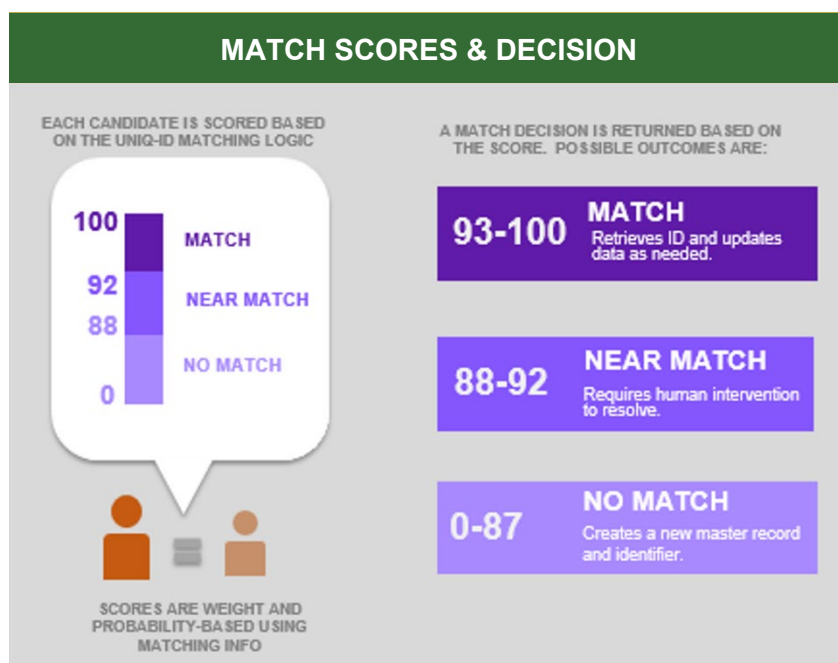
- The algorithm searches the master Student UID repository for existing records similar to the submitted record, based on predetermined matching fields. First name, last name, date of birth, and gender are required match fields; middle name, alternate last name, and social security number (SSN) also can be used for matching, if available.
- An index of potential matches is generated based on possible name combinations, as illustrated in Figure 4.2.

4.2 eScholar Uniq-ID Input, Index, and Search Flow



- The algorithm checks the potential matches for certain naming exceptions, such as errors in spelling, nicknames, multi-part names, uncommon names, and punctuation.
- Match scores are generated between the submitted record and similar records in the master repository.
- The match scores then are compared to the system's match score thresholds (which were established during the testing phase when the system was first used, with consultation with the vendor). Figure 4.3 outlines the three match score thresholds in the system.

4.3 eScholar Uniq-ID Match Score Thresholds



- **Match:** If the system finds an existing record that matches the submitted record with a match score of 93 or above, it designates that record as a “Match” and the UID of the matching student record automatically is assigned to the submitted student record, and the submitted data become part of the master record for that student. The previous master record (the information in the existing record *before* the addition of the submitted record’s information) is preserved in a UID record history.
- **No Match:** If the matching algorithm does not find a matching existing student record (match score below 87), a new UID is created and assigned to the submitted record.
- **Near Match:** If the system identifies one or more records in the master repository that have match scores that are too low to be considered a match but too high to be considered a “no match” (between 88 and 92), potential near match records are grouped for further resolution. Near match records must be reviewed and resolved by a system user before the system labels the submitted record as either matchable to an existing record or a new record—the system itself does not resolve near matches.

Once the assignment process is completed for all records in the submitted file, the records are updated in the master repository and an output file is generated that contains all of the submitted record information, a UID for each record, and an UID assignment resolution code for all submitted records.

5. North Carolina Entities that Assign UIDs⁵

Several NCLDS Contributors participate in the assignment of student UIDs. These Contributors include the North Carolina Department of Health and Human Services (DHHS) through its Early Childhood Integrated Data System (ECIDS) project, NCDPI, the North Carolina Community College System Office (NCCCS), the University of North Carolina System Office (UNCSSO), and the North Carolina Independent Colleges and Universities (NCICU). Each Contributor has developed its own process for submitting and assigning student UIDs. More than three million IDs have been assigned via the eScholar Student UID System since January 2022 alone. Table 5.1 provides an annual estimated number of assigned IDs for each Contributor, based on the past five years of submissions (2018 – 2022), rounded to the nearest one hundred thousand.

Table 5.1 – Estimated UID Assignment Breakdown

CONTRIBUTOR	ESTIMATED ANNUAL IDS ASSIGNED / UPDATED	PROPORTION OF TOTAL ASSIGNED IDS
ECIDS	400,000	13%
NCDPI	1,700,000	57%
NCCCS	500,000	17%
NCICU ⁶	—	—
UNCSSO	400,000	13%

ECIDS (DHHS)

Every month, ECIDS submits files of data records for children birth through age five who are receiving state services through such programs as Child Protective Services, Subsidized Child Care, various Pre-Kindergarten programs, Exceptional Children, Food and Nutritional Services, the Infant-Toddler Program, and Temporary Assistance for Needy Families.

NCDPI

Assigning eScholar UIDs to K-12 students takes place at the individual school level.

NCCCS

NCCCS uses the eScholar Student UID System to assign eScholar UIDs to all registered students (Curriculum, Continuing Education, and Basic Skills students).

⁵ Information in this section was obtained through direct communication with each Contributor.

⁶ Partner campuses represented by NCICU will not begin submitting postsecondary ICU student records for UID assignment and resolution until NCICU's proposed longitudinal data system becomes available.

NCICU

NCICU is not currently assigning UIDs in the eScholar Student UID System. It is anticipated that NCICU-affiliated institutions will resume UID assignment once NCICU's proposed new longitudinal data system is functional.

UNC System Office

UNCSCO assigns UIDs at the start of the Spring and Fall semesters.