IES Institute of Education Sciences

Planning for ECIDS Development: System Design Considerations

SLDS WEBINAR SUMMARY
August 2021

U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Statistics at IES

This product of the Institute of Education Sciences (IES) Statewide Longitudinal Data Systems (SLDS) Grant Program was developed with the help of knowledgeable staff from state education agencies and partner organizations. The content of this publication was derived from an SLDS monthly topical webinar that took place on August 26, 2021. The information presented does not necessarily represent the opinions of the IES SLDS Grant Program.

For more information on the IES SLDS Grant Program or for support with system development, please visit *http://nces.ed.gov/programs/SLDS*.

WEBINAR PRESENTERS
Mike Jolley and Stephen Matherly, *Early Childhood Utah*
Tanya Morgan and Hayley Young, *North Carolina Department of Health and Human Services*

MODERATORS
Amy Nicholas and Jeff Sellers, *SLDS Grant Program State Support Team*

Early childhood integrated data systems (ECIDSs) vary in many ways from state to state, from the legislative directives and regulations governing the system to their infrastructure and capacity. As the need to integrate and understand early childhood data increases, state agencies must consider not only initial requirements for the system, but also how the ECIDS might develop over time to meet new needs and challenges as they arise. Representatives from Utah and North Carolina share how they have successfully built their ECIDSs, including how their systems have evolved over time, key infrastructure decisions, and plans for future system enhancements.

Current Status of Early Childhood Data Integration

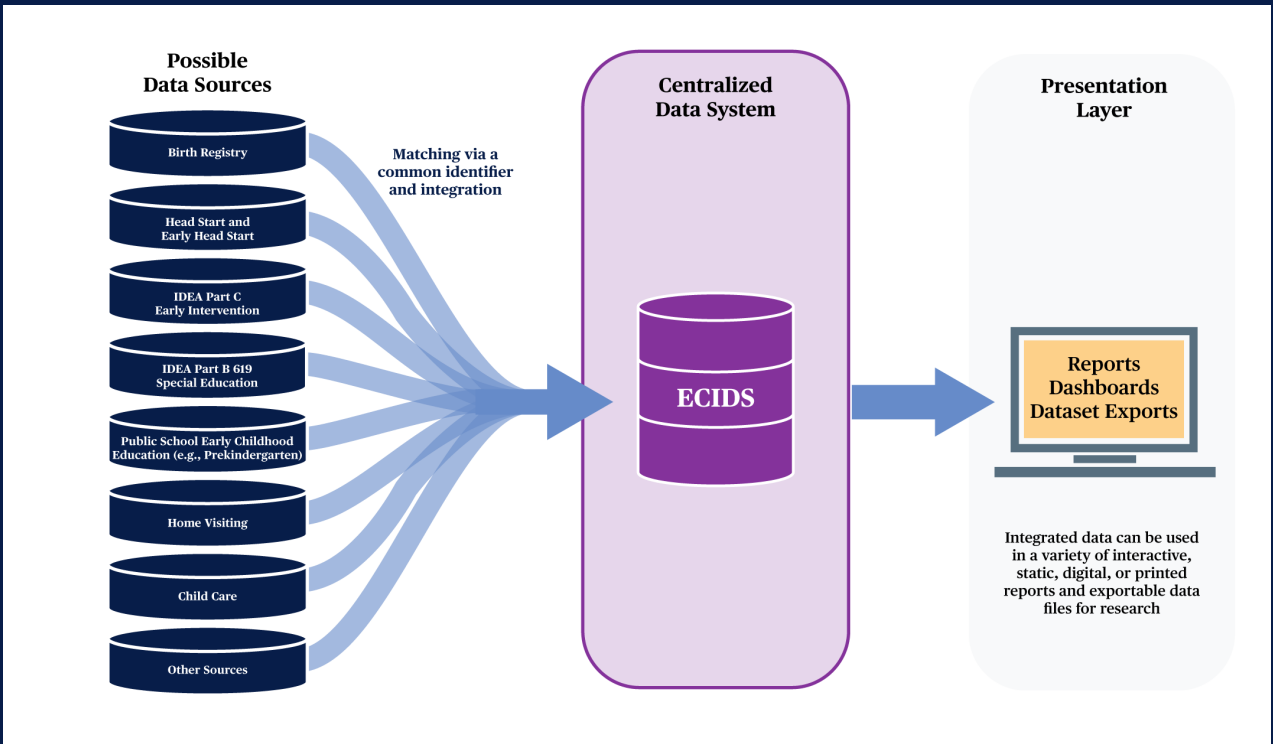
The Statewide Longitudinal Data Systems (SLDS) Grant Program conducts an annual survey to collect information from states and territories about the data elements, data linking methods, data uses, and infrastructure in their data systems. As of 2018, 39 percent of respondents said that they include early childhood data in a P-20W+ (early childhood through workforce) SLDS. Respondents also indicated that they house early childhood data in a separate, central early childhood system (43 percent of respondents) or in multiple early childhood data systems or source files (45 percent of respondents). Data from publicly funded prekindergarten, special education, and early intervention programs are most commonly linked in early childhood systems. The most common use of early childhood data is in state reports for the governor or legislature.

The survey also found that 31 percent of states and territories have a comprehensive data dictionary for early childhood data that includes data element names, definitions, option sets, field types, and formats.

ECIDS Technical Infrastructure Models

States use two main data models to maintain and provide secure access to data linked across organizations. In a *centralized* data system model, early childhood data from all participating programs and agencies are copied to a single, centrally located data repository where they are organized, integrated, and stored using a common data standard (**FIGURE 1** on page 2**)**.

FIGURE 1. In a centralized data system model, data from all participating programs and agencies are copied to a single, centrally located data repository.



Centralized models have the following strengths:

- Queries and reports can be run easily and in a timely manner.
- The ECIDS produces consistent data.
- It is easier to create a wide range of short- and long-term data reports.

Centralized models have the following limitations:

- The consolidated database requires extensive support, including a database administrator, storage capacity, and server.
- The public may be concerned about children's personally identifiable information being stored in one place or misused.

In a *federated* data system model, participating early childhood programs and agencies maintain control over their own data but agree through memoranda of understanding (MOUs) or data sharing agreements to share the data with other ECIDS partners upon request (**FIGURE 2** on page 3).

Federated models have the following strengths:

- There is no costly, centralized database to support.
- Federated models require fewer resources.
- There are fewer concerns about storing all child-level data in a central location.

Federated models have the following limitations:

- Determining longitudinal cohorts across data systems is challenging.
- The system can produce only limited-purpose data files; long-term and stored datasets are not available.
- The system cannot produce reports that rely on persistent data linkages.
- The linked data are not stored by the ECIDS but are cached, delivered to the requestor, and then removed.

Sometimes, states choose to create a *hybrid* model by picking features of the centralized and federated models to meet their unique needs (**FIGURE 3** on page 3). A hybrid ECIDS may maintain some linked data elements, such as identifiers, while destroying supplemental data when they are no longer needed.

Hybrid models have the following strengths:

- The matching process is done only once.
- Persisting data linkages cut down on processing time.

FIGURE 2. Participating early childhood programs and agencies in a federated model maintain control over their own data but agree through memoranda of understanding or data sharing agreements to share the data with other ECIDS partners upon request.

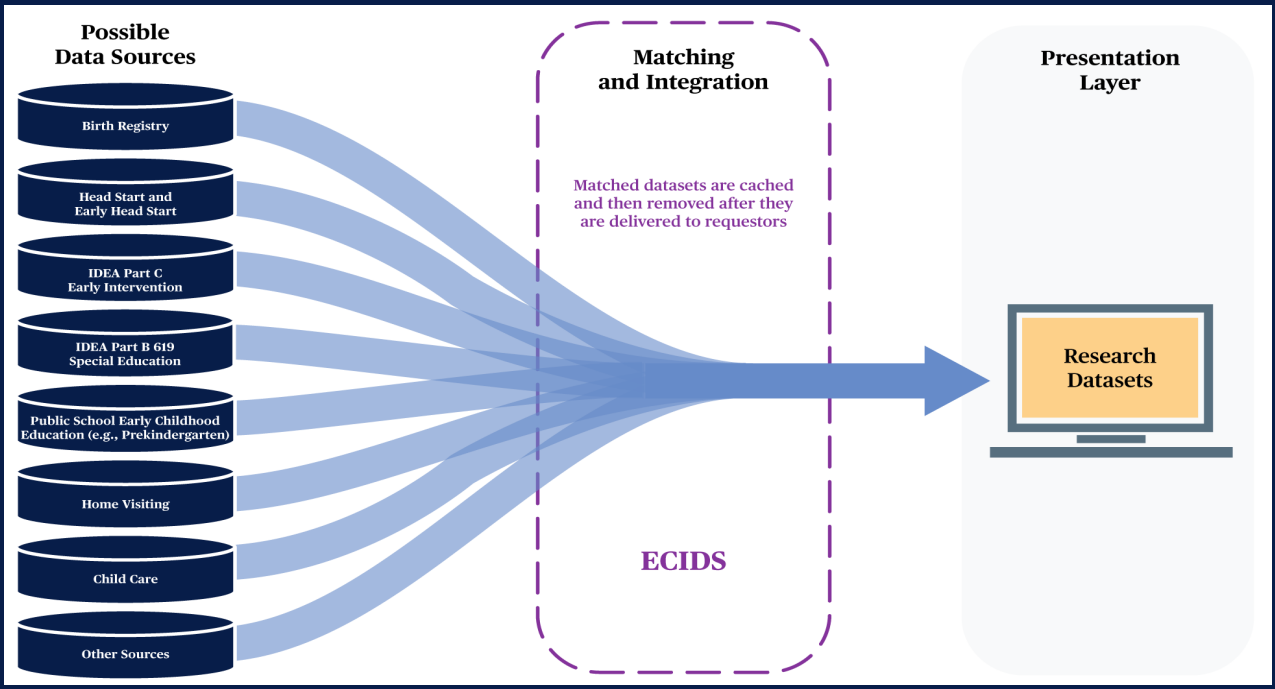
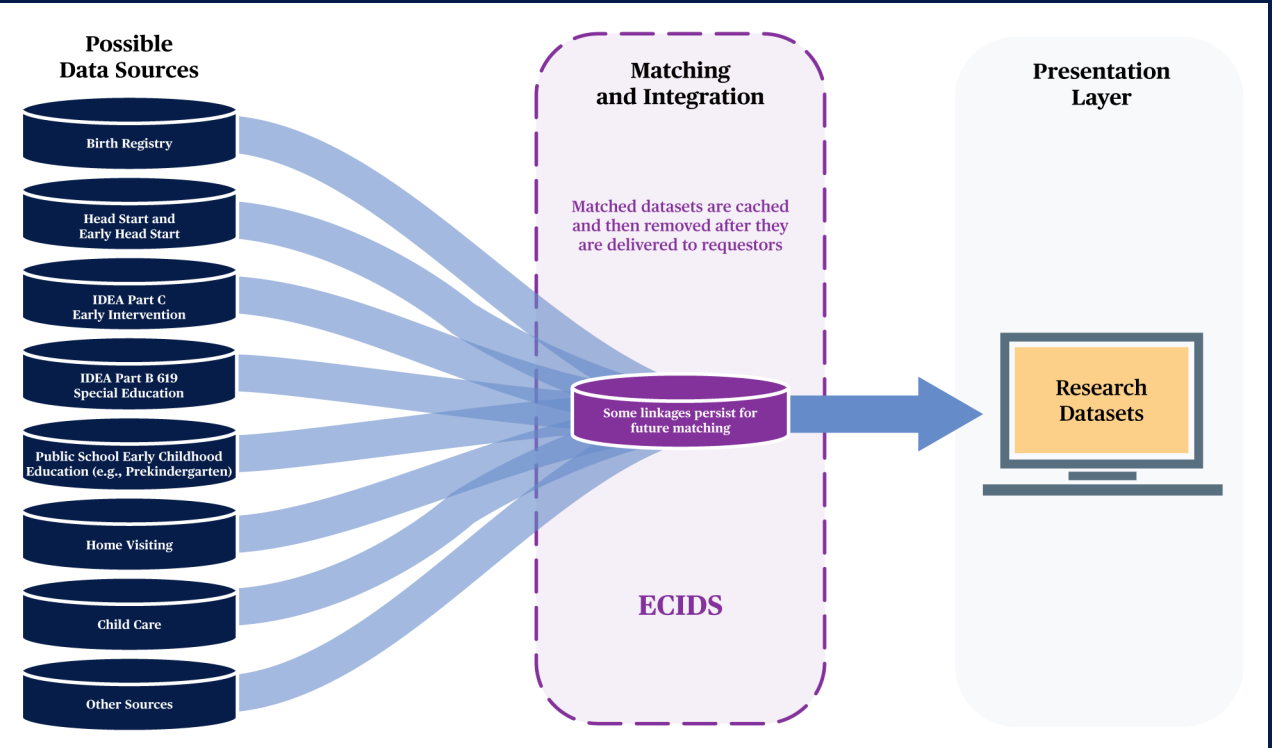


FIGURE 3. Hybrid models contains features of both the centralized and federated models.



- The system does not need a large central database or significant support to manage matched data.

Hybrid models often face similar reporting and cohort-defining challenges as a federated model.

Developing integrated systems for early childhood data requires four key steps. First, states should review their goals and vision to assess the needs, requirements, and objectives for the system. Next, they must design a system model that best meets their objectives. States will then develop the system architecture and data governance structure and identify the technology requirements for the system. Finally, they will implement the system and deliver data to stakeholders.

Utah's Early Childhood Integrated Data System

Basic considerations

When Early Childhood Utah first began planning its ECIDS, it examined ECIDSs in other states to determine how existing designs and technology might be leveraged to suit its needs. Staff members also inventoried other data systems in the state, including systems that would contribute data to the ECIDS. Understanding its partner agencies' data and systems helped Early Childhood Utah better determine ECIDS system requirements and stakeholder needs. Early Childhood Utah also explored whether the data sources already had assigned unique identifiers and how to match data across various early childhood systems and services.

In planning its ECIDS, Early Childhood Utah also considered how to engage stakeholders to ensure the system's sustainability as well as how to create documentation to answer key stakeholder questions, such as where data will be shared and how data will be protected.

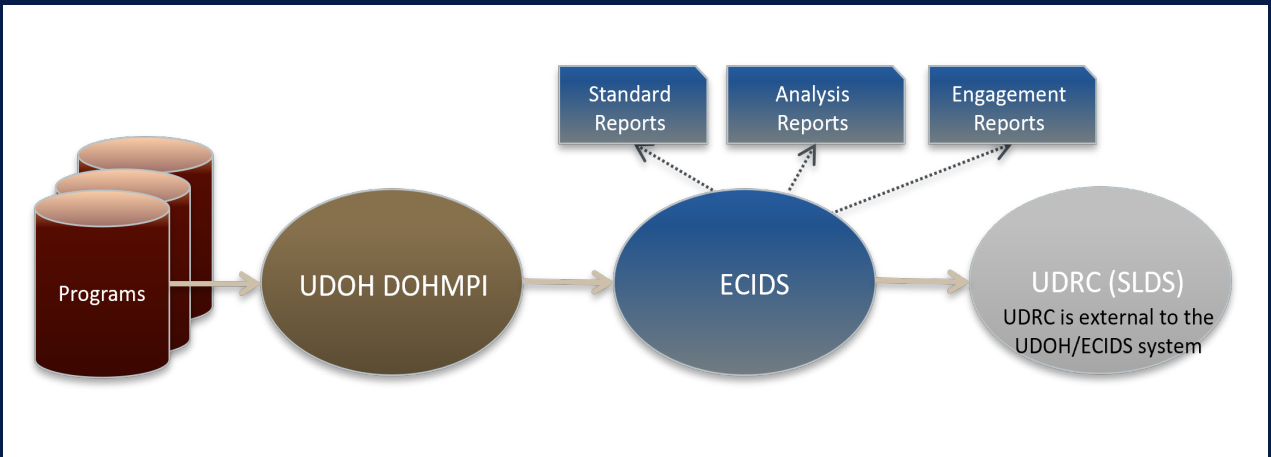
Long-term goals and objectives for the Utah ECIDS include

- ensuring confidentiality of protected data;
- minimizing impact on programs;
- using use cases to determine which program data to include in the ECIDS;
- creating report filters with flexible designs that meet stakeholders' overall goals and objectives;
- designing the ECIDS to be extended, maintained, and scaled in incremental steps;
- minimizing development and ongoing operational costs;
- using an iterative development lifecycle process; and
- maximizing opportunities for stakeholder and user engagement.

Infrastructure and workflows

Utah's ECIDS uses a system model that is not purely federated or centralized. It uses the Department of Health Master Person Index (DOHMPI) for extract, transform, and load (ETL) processes and identity

FIGURE 4. The Utah ECIDS relies on the Department of Health Master Person Index (DOHMPI) for data matching, extraction, translation, and exchange, while it provides data and research tools for longitudinal studies.



matching (**FIGURE 4** on page 4). Demographic data remain centralized in a different system (DOHMPI) from ECIDS. DOHMPI uses demographic data for matching, meaning that programs contributing data to the ECIDS do not have to maintain universal identifiers. Identifiable program data remain federated until they are de-identified and sent to ECIDS. ECIDS keeps de-identified data from programs in a central repository for use in research and reports.

The ECIDS receives data from sources such as the Individuals with Disabilities Education Act (IDEA) Part C Early Intervention program; some Head Start grantees; the Maternal, Infant, and Early Childhood Home Visiting program; Childcare Subsidy; the Women, Infants, and Children (WIC) program, and the Ages & Stages Questionnaire (ASQ) Enterprise Account, which hosts developmental and social-emotional screening data for children between ages 2 months and 5 years. The ECIDS also receives data from the statewide birth and death certificate registries.

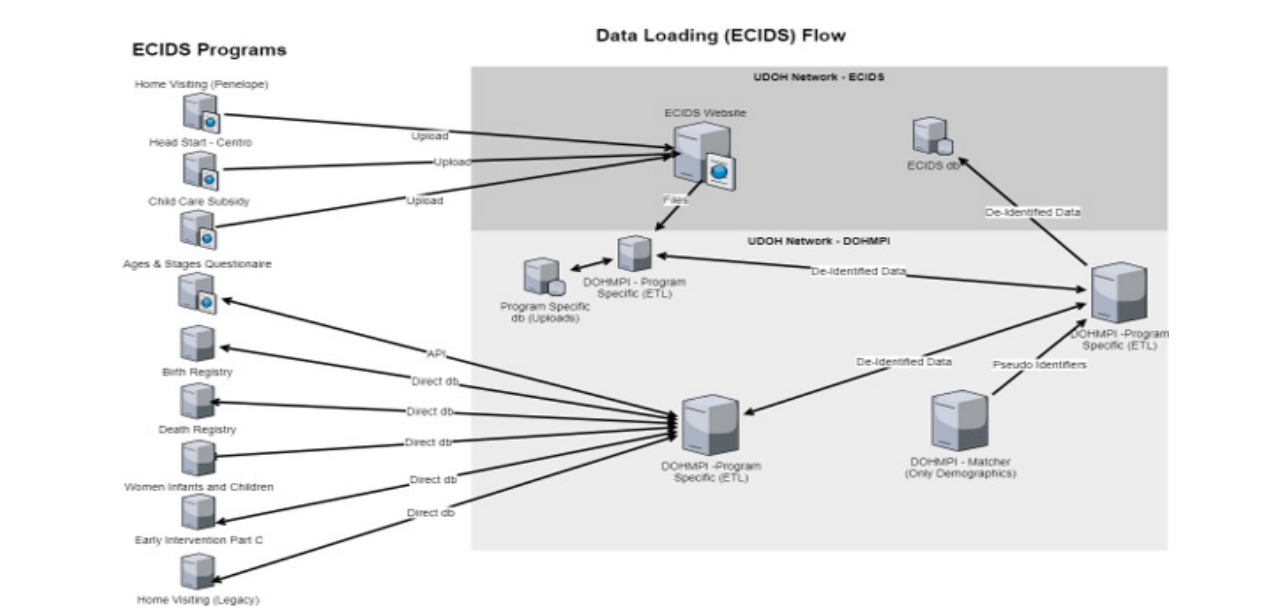
Data privacy concerns have played a role in Utah's ECIDS design. Users receive roles with specific access permissions and must access the system through assigned credentials. Keeping identifiable program data separate from the system has also helped Early Childhood Utah overcome many stakeholders' privacy concerns. Because matching is completed by DOHMPI rather than the ECIDS, ECIDS uses pseudo-identifiers, that make it more difficult to uniquely identify an individual within the ECIDS. The ECIDS also applies small cell size suppression to reports as applicable.

The ECIDS receives demographics data from its data loading matcher workflow. Programs such as Home Visiting, Head Start, and Childcare Subsidy upload data extracts to the ECIDS website, and their data are saved to program-specific databases. Other systems, such as the system that houses ASQ data, load their data through an application programming interfaces (APIs) or have their data directly integrated by using a direct database interface. A program-specific DOHMPI component completes the ETL process and sends the data to program-specific databases. Then, the component sends demographic data to a DOHMPI Matcher, which uses program specific identifiers and demographics such as names, dates of birth, and addresses that are common across early childhood programs. The DOHMPI Matcher then sends pseudo-identifiers alongside the de-identified data to the ECIDS (**FIGURE 5**).

Utah ECIDS reports can be filtered by zip code, county, age, gender, race/ethnicity, and dynamic timeframes. Access to ECIDS reports is governed by the ECIDS data sources. Utah's ECIDS currently provides the following data products:

- The Early Childhood Utah Community Assessment Tool (CAT)
    - Population Eligibility: Child population data and household composition

FIGURE 5. The DOHMPI Matcher sends pseudo-identifiers that are used for the ECIDS's data.



    - Risk Factors: Child population by age, child risk factors, living in poverty, and mother's child risk factors at birth
    - Provider Counts: Provider counts by service category and region
    - Capacity: Child-care slots and vacancy by age group and region
    - Resources: Links to Utah's Preschool Entry and Exit Profile (PEEP), Utah's Kindergarten Entry and Exit Profile (KEEP), and Utah Education Scorecard
- Standard Reports
    - Distinct child counts within a program and child counts across programs
    - Children enrolled or not enrolled, children in single and multiple programs, and newly or previously enrolled
    - Distinct counts of children being served in multiple programs
    - Child program enrollment sequence patterns
- ASQ Screening Reports
    - Overall and program-specific ASQ screening counts and overall results by age intervals
    - Distinct children screened by multiple ASQ programs
    - Child counts in single and multiple ASQ programs

Challenges and lessons learned

Data governance is key for an effective ECIDS. Stakeholders change over time, and while one agency representative may be excited about sharing data, their replacement may have more privacy concerns. Early Childhood Utah recommends that ECIDS staff members thoroughly document procedures, workflows, data governance manuals, and data sharing agreements and MOUs so that they can quickly onboard new stakeholders.

Differences in how data source systems collect child-level data may present challenges for an ECIDS. Some source systems collect only family-level data, but the ECIDS needs child-level data to correctly match individuals with one or more services. Data granularity can also be a problem, as different agencies may collect less detailed data on information such as race and ethnicity.

Next steps

Early Childhood Utah is currently in the process of making limited ECIDS data publicly available through its Community Assessment Tool. The ECIDS is also integrating program engagement and attendance data and developing related reports that will display program engagement, class attendance, and/or home visiting data. Data of this nature are critical to performing longitudinal research into the impact of early childhood intervention on K-3 outcomes and beyond. Early Childhood Utah anticipates sharing ECIDS data with Utah's SLDS, the Utah Data Research Center.

North Carolina's Early Childhood Integrated Data System

North Carolina's ECIDS (NC ECIDS), which is maintained by the North Carolina Department of Health and Human Services (NCDHHS), is the only system in the state that integrates data from early education, health, and social services programs to inform key program and policy decisions about children ages 0-5 and their families. The system was originally funded by a Race to the Top - Early Learning Challenge grant. The North Carolina Department of Information Technology manages technology for NC ECIDS. The NC ECIDS Governance Committee sets policies for NC ECIDS and provides feedback to inform development of its data application, web portal, and workflow processes. NC ECIDS contains data only for children who receive services from participating programs, rather than all children in North Carolina. NC ECIDS data are used to

- answer key program and policy questions;
- provide counts of children who receive multiple early childhood services;
- produce aggregate reports for common program and policy questions; and
- provide data elements for approved research requests.

Data, infrastructure, and reports

NC ECIDS currently

- integrates early childhood data across programs;
- provides distinct counts of children served;
- links data such as from IDEA Part C and Part B for reporting; and
- allows researchers to request data to answer key policy and program questions.

NC ECIDS receives data from education, health, and social services programs, including

- North Carolina prekindergarten;
- Subsidized Child Care;
- Infant Toddler Program - Early Intervention (IDEA Part C);
- Special Education (IDEA Part B 619);
- Food & Nutrition Services;
- Child Protective Services; and
- Temporary Assistance for Needy Families (TANF).

Data from these programs are matched using a unique identifier through the eScholar platform (**FIGURE 6**).

NC ECIDS staff members are collaborating with the system's program and technology partners to add updated data to existing NC ECIDS reports, which currently have only data through 2014-2015 publicly available. Staff members are also developing a new NC ECIDS website.

NC ECIDS currently produces four reports:

- Total and Unduplicated Number by NC ECIDS Service, which provides information on how many children used North Carolina early childhood services during recent fiscal years
- Number of Children Receiving Multiple NC ECIDS Services, which reports how many children used two or more programs during fiscal years
- Number of Children Receiving Two NC ECIDS Services, which shows how many children attended a combination of two programs during recent fiscal years
- Number of NC Pre-K Age Eligible Children Receiving NC ECIDS Services, which illustrates how many North Carolina prekindergarten-age eligible children are receiving NC ECIDS services

Next steps

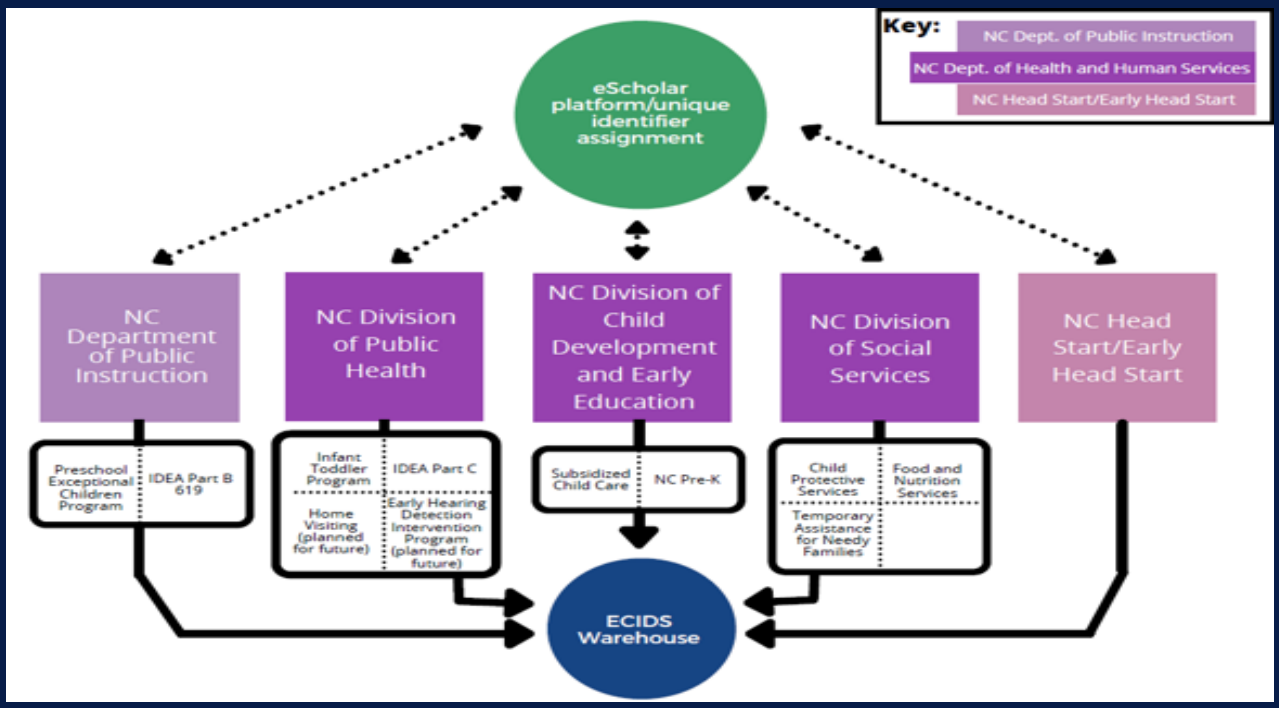
NCDHHS is developing a data selector tool that will expand functionality for submitting, processing, and managing internal and external data requests. Users will be able to see specific program data by clicking on a tab and can filter data by fiscal year, county, or certain demographics such as gender. NC ECIDS staff members will also work with North Carolina's SLDS team to incorporate early childhood data into the state SLDS.

North Carolina recently received funding from the Preschool Development Grant to modernize and expand NC ECIDS, resulting in expanded reports and reporting functionality for internal NC ECIDS teams and the public. Data from Head Start, home visiting programs, and North Carolina's Early Hearing Detection and Intervention program will be integrated into NC ECIDS, as well. The sustainability and cost efficiency these additions provide are a priority for the NC ECIDS team.

Additional Resources

Early Childhood Utah
https://earlychildhoodutah.utah.gov/

FIGURE 6. After programs upload their data, the eScholar platform assigns unique identifiers to an individual's records and sends the data to the NC ECIDS data warehouse.



North Carolina Division of Child Development and Early Education
https://ncchildcare.ncdhhs.gov/

North Carolina Early Childhood Integrated Data System
https://www.ncdhhs.gov/north-carolina-early-childhood-integrated-data-system

SLDS Issue Brief: Centralized vs. Federated: System Models for P-20W+ Data Systems
https://slds.ed.gov/#communities/pdc/documents/17542

SLDSSLDS Early Childhood Integrated Data Systems Toolkit
https://slds.ed.gov/#program/ecids-toolkit

Utah Early Childhood Integrated Data System
https://ecids.utah.gov/