The Early Childhood DATA Collaborative

Developing Coordinated Longitudinal Early Childhood Data Systems

Trends and Opportunities in Race to the Top Early Learning Challenge Applications

September 2012

State policymakers are increasingly focused on the importance of quality early care and education (ECE) experiences in closing the achievement gap and preparing all children to succeed in school and in life. Data are a vital resource that can inform state ECE policies and investments, yet access to timely, reliable data on children before they enter kindergarten is scarce. Most state ECE data systems function primarily to satisfy reporting requirements for different federal programs, making information on young children's ECE experiences uncoordinated, potentially duplicative, and not necessarily responsive to state needs. As a result, policymakers cannot get answers to basic questions, such as how many children currently participate in high-quality ECE programs? How many more could benefit if they had access? Are these programs improving child outcomes?

The Race to the Top Early Learning Challenge (ELC) encouraged states to demonstrate their commitment to integrating and aligning resources and policies across all of the state agencies that administer public funds related to early learning and development. Building or enhancing an early learning data system was an optional section of the application, and 30 of the 37 applicants addressed this priority. In 2011, nine states, six of which addressed the data priority, won competitive ELC grants.[1] An additional five states are eligible to apply for smaller grants in 2012, and all but one of these states addressed the data priority in their initial applications.[2]

Some states that chose not to address this priority in their ELC applications indicated that they were already working on early learning data systems through other federal grants, including the State Longitudinal Data Systems (SLDS) Grant or grants that support State Advisory Councils on Early Childhood Development and Care.

This issue brief analyzes the 30 state plans to build or enhance early learning data systems (section E2), with a more in-depth review of the states with the top 17 scores from this section. It reflects only what states described in their ELC proposals, not what has been or will be implemented since many states did not receive ELC grants. It is not a comprehensive report of what states are doing, but a window into what states are thinking about and a guide for future trends in the development of integrated state ECE data systems. These trends include the following:

- **making data accessible to improve and inform ECE practice and policy**
- **linking existing ECE data systems**
- **filling ECE data gaps, including workforce and child development data**
- **strengthening the connection between ECE data and data from other systems**
- **developing interagency data governance structures**

Making Data Accessible to Improve and Inform ECE Practice and Policy

The success of a well-designed and functioning ECE data system rests on its usefulness. Does it answer critical policy questions? Does it help program administrators improve program quality, early learning providers improve instruction, and parents understand what their children know and can do? Many states used the ELC grant opportunity to articulate a plan for an integrated data system that would make data accessible to varied stakeholders to support continuous improvement. For example, **Massachusetts** set the following goals for its data system:

- providing policymakers with information about the current use of early learning/development programs that could be disaggregated by local level/different groupings of children to better address gaps in services and effective practice
- providing programs and services with information about the children they are serving to improve individualized teaching and learning
- providing an opportunity for state agencies to understand where children may be served by multiple systems that would benefit from greater coordination and integration
- providing parents/families with information about the early learning/development programs that are available to them and information to support their children's development

In their ELC applications, states proposed new data portals, dashboards, scorecards, and reports tailored to specific users, all available online. They pledged to make better use of technology to put data into the hands of people who could use them. For example

- **Maryland** and **Florida** proposed web portals to streamline and improve the timeliness, completeness, and accuracy of system records related to the providers and consumers of child care services.
- **Pennsylvania** proposed developing a "provider scorecard" that would compile data on individual ECE program sites, including provider demographics, Quality Rating Implementation System (QRIS) information, technical assistance, classroom quality assessment scores, workforce qualifications and retention, child demographics, enrollment, and child outcome information. Providers would receive comprehensive training on the effective use of their scorecard and recommendations for program improvement.

An Important Caveat: Variations in How States Defined the Data Landscape

This issue brief highlights themes across ELC applications, but significant variations in how states defined the scope of their data landscape and the data elements would affect how data could be used.

Targeted vs. Universal: Most states focused on integrating data from one or more publicly funded early childhood programs, such as subsidized child care, Head Start, state prekindergarten, early intervention, preschool special education, and home visiting. In these states, data use would be limited to the children and programs receiving public funds. However, a few states proposed a larger scope for their early learning data systems. **Rhode Island** proposed a universal child-level database, starting at birth, building on the public health data system that tracks birth record, immunization, and newborn screening data for virtually all young children in the state, linking the public health data system to the K–12 SLDS. **New York** and **Maryland** wanted to include all children in ECE programs, not just those receiving publicly funded subsidies.

Voluntary vs. Mandatory: Several rich sources of state ECE data rely on voluntary participation or require participation from only select groups of program sites and providers. These include the state Tiered Quality Rating Implementation System (TQRIS), which rates the quality of ECE program sites based on an array of program characteristics, and state workforce registries, which collect detailed information on the qualifications of the ECE workforce. While several states proposed using ELC grants to expand participation in both TQRIS and workforce registries, the information in these databases most often does not represent the full data element.

Datasets that rely on voluntary reporting or that represent only a portion of the population limit the usefulness of the data in decision making.

- **Minnesota** proposed the creation of a new portal to provide user-friendly, web-based dashboards and reports tailored to educators, administrators, and parents. This would include reports that could link individual teachers and students and could link child attendance and student assessment data.
- The **Connecticut** Data Collaborative is an open affiliation of individuals and organizations from the public and private sectors interested in improving ways data can be accessed and used. The collaborative currently has an Early Childhood Portal that provides access to a wide range of early childhood indicators such as birth and death rates, prenatal care, prekindergarten exposure, school test results, and participation in ECE programs. In the ELC application, Connecticut proposed disseminating data to local early childhood councils and the public through an open-source, web-based platform with enhanced data analysis and data visualization features to show time trends and compare data across geographic areas.

States described multiple ways to make data more accessible with the goal of making them more useful, but the ultimate usefulness of the data depends on the quality of the data collected and the ability of users to interpret the data. Many states would limit their data collection to a target group of programs or rely on voluntary reporting, significantly limiting the usefulness of the data. (See "An Important Caveat: Variations in How States Defined the Data Landscape.") Several states emphasized the importance of building users' capacity to understand and use data. **Rhode Island,** for example, proposed incorporating effective provider data practices into both QRIS and preschool program standards.

Linking Existing ECE Data Systems

In their plans to develop coordinated early childhood data systems that would provide useful information to policymakers and practitioners, states faced the daunting challenge of linking existing ECE databases across program and agency silos and across levels of data. States proposed two different approaches: data warehouses and federated data systems. **Pennsylvania** and **Maryland,** for example, each proposed an interagency ECE data warehouse—a central hub to house data from different agencies. Other states, such as **Connecticut** and **New Mexico,** proposed federated data systems in which data would remain in existing agency databases, but a user would be able to extract and analyze data across program and agency silos.

Data experts report that a federated system can be less costly to develop and can simplify the process of complying with legal requirements around disclosing client data. A data warehouse, however, more easily facilitates large-scale research and policy analysis across the various programs serving young children.[3]

Regardless of which approach a state chose, state leaders identified several important strategies to link ECE data. These strategies included the following:

- **identifiers** to accurately match records among datasets that represent the same child, program site, or provider/teacher
- **common data standards** to ensure that data fields represent the same type of information when linking databases
- **data-sharing agreements** to develop formal documents that define how data would be linked and used

Identifiers

Assigning a single, nonduplicated identifier to each child, ECE program site, and member of the ECE workforce facilitates matching records across databases. The Early Childhood Data Collaborative (ECDC) 2011 survey found that only one state **(Pennsylvania)** could link appropriate child- and program site-level data across all ECE programs, and no state could link individual ECE workforce-level data across ECE programs.[4] (See box on page 4 for more information about the survey.)

Several ELC applicants proposed linking child-level records by assigning existing identifiers, such as birth certificate numbers **(Connecticut)** or the K–12 secure student identifier. **New York** assigns the K–12 identifier upon entry into the state prekindergarten or preschool special education programs and proposed expanding this practice to other programs, such as child care subsidy and early intervention. Other states (such as **Maine**) did not plan to use unique identifiers but instead proposed matching existing data fields (first name, last name, address, birth date) to identify and link the same child, program, or teacher across databases.

Some states are using both of these approaches. For example, **Illinois** currently assigns the K–12 student identifier to children in a number of early learning programs (including prekindergarten and birth–3 programs). At the same time, the state has launched a "Common Identifier" project to facilitate accurate matching of existing child identifiers across a wider range of state programs serving children. A single, nonduplicated identifier yields the highest matching rate when linking separate datasets, but implementing a common identifier can be costly, time consuming, and politically unpopular. If a state is largely interested in using data for population-based analysis, then state leaders report that a matching algorithm may meet their needs.

Common data standards

A data standard is a set of commonly agreed-upon names, definitions, option sets, and technical specifications for a given selection of data elements. For example, if the data field contains enrollment information, the definition may specify actual enrollment data at a point in time or cumulative enrollment.

Before data fields can be linked across systems, states need to ensure that the fields refer to the same concepts, or else any information drawn from the data will not be valid or meaningful. To do so, several states proposed implementing common data fields and data definitions across existing systems. **Connecticut** proposed offering small grants to allow agencies to modify existing databases to ensure standardized state data structures, formats, and definitions. Other states, including **Minnesota, Massachusetts,** and **North Carolina,** discussed plans to adopt the early childhood component of the Common Education Data Standards.[5] This common set of data standards, supported by the U.S. Department of Education and developed by a consortium of education stakeholders, provides a structure that any state can voluntarily use to align data fields.

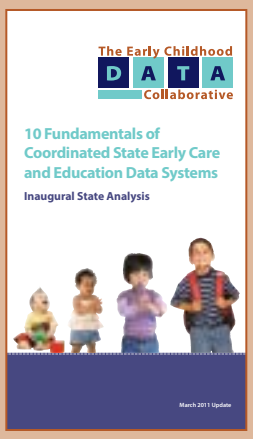
Data-sharing agreements

Data-sharing agreements address the policy and legal challenges associated with linking data, such as how the data will be used; who will have access; and how to ensure the privacy, security, and confidentiality of personally identifiable data. Many states, including **Maine, New York,** and **Illinois,** proposed formal memorandum of understanding agreements (MOUs) to allow data sharing across agencies. **New York,** for example, proposed that eight state and two New York City agencies sign MOUs to formally establish a coordinated data system.

States Collect ECE Data but Cannot Transform the Data into Actionable Information

In March 2011, the Early Childhood Data Collaborative (ECDC) released findings on how states are collecting and linking data on children, the workforce, and program sites in six state ECE programs: subsidized child care, licensed child care, state-funded Head Start/Early Head Start, state-funded prekindergarten, early intervention, and preschool special education. The analysis revealed the following:

- **Every state collects ECE data** on individual children, program sites, and/or members of the ECE workforce for at least some of the state's ECE programs.
- **Data gaps remain,** especially in workforce data and data related to child development.
- **Data are uncoordinated,** as only one state could link data across all ECE programs at the child and site levels, and no state could do it at the workforce level.
- **Governance matters** because data linkages are more likely to occur between data systems located within the same state agency.



The ECDC formed in 2009 to promote policies that support states' development and use of coordinated ECE data systems. By articulating the Ten Fundamentals of Coordinated State ECE Data Systems and monitoring state progress in developing coordinated ECE data systems, the ECDC seeks to support data-driven policymaking. When data are accessible and stakeholders have the capacity to use data appropriately, coordinated ECE data systems can provide answers about how to improve the quality of ECE programs and the workforce, increase program access, and ultimately improve child outcomes.

For more information on the ECDC, the Ten Fundamentals, and state-by-state results of the analysis, go to **www.ECEData.org**.

Filling ECE Data Gaps

While the majority of state ELC plans reflected an intentional effort to make better use of existing data on young children and ECE programs, states also proposed filling some critical data gaps. First, states proposed collecting new data elements required by the ELC application process. These elements included program-level data on the program's structure, child suspension and expulsion rates, staff retention, staff compensation, and work environment, as well as data reported as part of a TQRIS. The ELC grant application also required states to collect child-level program participation and attendance data.

In addition to addressing these grant requirements, state applications focused on filling two significant gaps in data collection:

- data on the ECE workforce
- data related to children's development over time

Expanding data on the ECE workforce

The 2011 ECDC state analysis found that states collect less data on the workforce than on programs and the children they serve, and many of the ELC applications focused on filling this data gap.[6] State workforce registries are information systems that capture data about the early childhood workforce and often serve as a hub for information on training opportunities available in the community. As of 2012, 33 states had state registries, and several other states were planning or developing registries.[7] Some states require practitioners in certain sectors to register; in other states, registries are voluntary. State ELC applications reflected an interest in making these registries more robust, by adding data fields and increasing participation in the registry.

State registries typically capture demographic, educational, professional development, and employment information on individuals working in licensed child care. **Connecticut,** for example, proposed new data fields that would focus on whether practitioners meet the requirements specified in the state's core knowledge and competencies framework. Other states, such as **Vermont,** are standardizing data elements so they comply with the National Registry Alliance standards. By doing so, these states will have the capacity to link workforce data with other states.

Many states focused on expanding their registries to more accurately reflect the state early childhood workforce, which often extends beyond center-based teachers and home providers serving children from birth to age 5. **Illinois** and **Vermont** proposed linking early childhood workforce registries to data on certified teachers working with children from birth to age 8. **Illinois** hoped to increase participation in the registry from 18,500 to 50,000 by requiring all practitioners working in licensed early learning development programs to register. **Vermont** also proposed outreach to practitioners employed outside licensed ECE programs, such as health workers and early interventionists.

Collecting data on child development

Access to child development data can inform improvements in instruction, promote smooth transitions between programs, and potentially reduce duplicate assessments when children move from one early childhood program to another. The early childhood field, however, has not yet reached a consensus around when and how to effectively assess child development. As a result, state leaders have been reluctant to mandate specific assessment protocols. While the ELC grant opportunity has spurred new state proposals around assessment, any state effort to collect this information will need to include significant stakeholder input on appropriate assessment tools and uses of assessment data.

Many ELC applications mentioned the desire to capture new data on child development, ranging from developmental screenings to kindergarten entry assessments (KEAs), in ECE data systems. For example

- **Rhode Island** planned to expand KIDSNET, the data system that houses public health data such as birth records, immunizations, and newborn screening data on all children born or immunized in the state. KIDSNET would be expanded to serve as a central, statewide depository for developmental screening and referral data for children from birth to kindergarten entry and would link to early learning data and the K–12 SLDS.
- **New Mexico's** Community Data Collaborative makes aggregate early childhood health data available at the community level. The state application proposed expanding this work by adding assessment data from prekindergarten and child care programs participating in the state's TQRIS into the community-level data-mapping tool.

- **Maryland** described a process to collect and report on child formative assessment information across publicly funded programs participating in the state's revised TQRIS.
- **Thirty-five of the 37 states** applying for an ELC grant (all applicants except Maine and Missouri) proposed developing, revising, or expanding a statewide KEA that would cover five required domains of development and align with state early learning standards. Twenty of these states proposed creating a new statewide KEA, while others would revise and/or scale up an existing assessment. As required by the ELC application, these states all would include KEA data in their integrated data systems, and many also proposed specific new uses of KEA data to inform kindergarten instruction, as well as ECE policy and practice.

States are required to report on child outcomes for children participating in programs funded through the Early Intervention (Part C) and Preschool Special Education (Part B, Section 619) funding streams. In general, the applications reflected how states plan to or are already linking these data to their K–12 SLDS.

Strengthening the Connection between ECE Data and Data from Other Systems

The ELC application required all states to integrate or link ECE data to an SLDS. As a result, several states proposed new strategies to connect ECE data with their existing SLDS. Several states also went beyond the requirements, proposing links to other data systems, such as federally funded Head Start and health and human services data systems.

Linking with K–12 data systems

When ECE data are linked with an SLDS, teachers and administrators can track and support individual children's educational progress from ECE programs through elementary school and beyond. With these data linkages in place, policymakers and researchers may also be able to analyze long-term outcomes for children who participate in publicly funded ECE programs.

Several states proposed new strategies to link child-level data between these two data systems. For example, **Pennsylvania** proposed linking existing unique ECE identifiers with K–12 student identifiers across the state's ECE and K–12 data warehouses. These links would allow reporting on 3rd grade outcomes for children served in state ECE programs. In **Connecticut,** which is using birth certificate numbers as a common identifier for children in ECE programs, the state proposed linking this identifier with the state's K–12 student identifier upon children's registration in kindergarten.

Other states proposed building one integrated P–12 or P–20 education data system. **Illinois,** for example, is using a federated data integration model for a P–20 education data system and assigning unique child identifiers across multiple ECE programs. This system has expanded to include new ECE programs in recent years, including prekindergarten, home visiting, and early intervention.

Integrating Head Start

At least nine states proposed integrating federally funded Head Start and Early Head Start data into a coordinated state early childhood data system. This integration would be particularly challenging since Head Start data reside with local grantees. While aggregated data are reported to the federal government, there is no state repository for data on children participating in Head Start. According to the applications, **Maine** and **Illinois** are working to collect federally funded Head Start data at the state level. **Maine** is integrating data from Head Start and Early Head Start grantees into its SLDS. The Illinois Head Start Association has a plan to develop a state-level Head Start database containing child-level demographic and developmental data and program site (grantee) information for all Head Start and Early Head Start programs in the state. Head Start and Early Head Start workforce data would be captured through the states' workforce registries.

Linking with health and human services data systems

Finally, several states proposed linking ECE data to related data systems, such as health and human services data. **New Mexico** mentioned integrating child welfare and the Child and Adult Care Food Program data. **Vermont** proposed linking to the Medicaid Management Information System for developmental screening and well-child visit information. **Pennsylvania, Illinois,** and **Maine** proposed integrating data from one or more home-visiting programs.

Developing Interagency Data Governance Structures

Many states proposed developing a state data governance body to set state policies that would guide data collection, access, and use. These policies would ensure data quality; protect privacy, security, and confidentiality; and ensure interoperability between new and existing state data systems. Since ECE data typically exist in multiple state agencies, data governance bodies need to reflect the interagency nature of the data system. It is critical to include several different types of individuals in this body, including decision makers with authority over budgets and agency policies, stakeholders who understand the meaning behind the data and how data would be used, and information technology or data managers who understand technology systems and privacy and security safeguards.

While some states, such as **Connecticut,** proposed new interagency early childhood data governance bodies, most states identified the state department of education as the lead agency and proposed adding early childhood data to existing governance structures. **Maine** and **Kentucky** proposed expanding the existing SLDS steering committee. **Rhode Island, North Carolina,** and **Vermont** proposed establishing a data system governance body within the existing Early Childhood Advisory Council (ECAC) structure, which already serves as a hub for interagency collaboration and includes stakeholders from state and local agencies and programs serving young children. **North Carolina** also proposed a process to engage local Smart Start and other local coordinating councils in providing input on plans to collect and use ECE data. **Maryland** proposed leveraging the expertise of both its SLDS committee and its ECAC. The SLDS committee would serve as the governance body to support improvements to the state's Early Childhood Data Warehouse, and the ECAC would provide assistance and input to this committee to ensure that the data warehouse meets stakeholders' data needs.

Looking Forward

Through the ELC, states were encouraged to redesign the use of their data systems, changing the focus from compliance to continuous improvement. States that won the ELC competition now have the opportunity to make better use of existing data and develop a data infrastructure that can support a high-quality early childhood system. Most applicants, however, did not receive an ELC award and reported only limited efforts to carry out their plans with support from other federal grants, state funds, or private funds.

Given continued challenging fiscal climates in states, states that did not win ELC grants are moving forward with lower-cost actions in their plans (e.g., mapping existing data elements) that will build the groundwork for a statewide ECE data system. Early childhood stakeholders in some states are reaching out to partners in K–12 or health systems that may have resources and expertise on integrated data system development. Finally, some state leaders are promoting incremental steps, such as assigning a K–12 student identifier to children in targeted ECE programs, piloting coordinated data efforts in a local community, or creating a prototype for an integrated data system.

As state leaders implement some or all of their ELC data systems plans, they will benefit from opportunities to learn from each other. While state plans reflect different approaches to building coordinated ECE data systems, there are many commonalities and opportunities to share promising strategies. Specifically, states need assistance in understanding best practices in sharing information with various stakeholders through user-friendly web portals and reports and linking existing databases. States will also benefit from continued sharing of best practices in strengthening workforce registries and child assessment data, linking ECE data with other state data systems, and developing effective strategies for data governance. As they implement their plans, states that won ELC grants will offer new lessons and provide examples of the potential of integrated data to inform policy and practice.

Endnotes

1 The nine states that were awarded grants are California, Delaware, Maryland, Massachusetts, Minnesota, North Carolina, Ohio, Rhode Island, and Washington. California, Delaware, and Washington did not address the data systems priority in their applications.
2 These states are Colorado, Illinois, New Mexico, Oregon, and Wisconsin.
3 http://www.ispc.upenn.edu/documents/Prashant.pdf
4 ECDC survey, November 2010
5 https://ceds.ed.gov/
6 ECDC Survey, November 2010
7 http://www.registryalliance.org

ACKNOWLEDGMENT

This analysis was produced by members of the ECDC, with special thanks to Fran Kipnis, Center for the Study of the Child Care Workforce; Helene Stebbins, Birth to Five Policy Alliance; and Amanda Szekely, National Governors Association Center for Best Practices, who authored the brief. The ECDC gratefully acknowledges the contributions of the many national partners who helped analyze the Race to the Top Early Learning Challenge grants and our partners at the state level who reviewed the contents of this brief.

The Early Childhood DATA Collaborative

A partnership of The Center for the Study of Child Care Employment at UC Berkeley, Council of Chief State School Officers, Data Quality Campaign, National Conference of State Legislatures, National Governors Association Center for Best Practices, and Pew Home Visiting Campaign. Child Trends serves as the central hub for the ECDC.

The Early Childhood Data Collaborative (ECDC) supports state policymakers' development and use of coordinated state early care and education (ECE) data systems to improve the quality of ECE programs and the workforce, increase access to high-quality ECE programs, and ultimately improve child outcomes. The ECDC will provide tools and resources to encourage state policy change and provide a national forum to support the development and use of coordinated state ECE data systems.

The ECDC is supported through funding from the Birth to Five Policy Alliance, The David and Lucile Packard Foundation, and an anonymous donor.

For more information, please visit www.ECEData.org.