

## Policy Brief

# Improving Data Quality through Data Governance

**Analysts:** Tim Bedeaux, Senior Policy Analyst II

**Date:** November 14, 2024

High-quality data serves as the foundation for strong analyses of educational programs. The Legislature relies on data from the Public Education Department (PED) and other state agencies to evaluate whether policies, programs, and funding meet the diverse needs of students across the state. For policymakers, access to reliable and accurate data is essential in ensuring the budget and statutory programs align with the state's long-term educational goals.

New Mexico faces significant data integrity challenges; often, datasets are rife with issues in accuracy, completeness, consistency, granularity, and timeliness. Issues in the quality of data have impacted the state's ability to implement and evaluate educational programs. Throughout the 2024 legislative interim, poor data quality has challenged PED's ability to implement programs and identify areas for improvement, and further, has complicated LESC's efforts to assess program impacts.

The root of New Mexico's data issues lies in the state's lack of strategic oversight and alignment among the state's many data systems. PED has created data systems primarily to meet its own operational needs, but these systems are not always aligned with the needs of policymakers, analysts, educators, and families, who rely on data for different purposes.

PED would benefit from a formal structure for data governance, including rules, policies, and administrative oversight designed to ensure the state's educational data are secure, private, accurate, available, and usable, and comparable over time. This structure may also include additional membership from other state agencies; other states have begun to establish interagency data governance boards to improve the quality of cross-agency data sharing and interoperability of education data systems.

This brief explores the challenges of New Mexico's current data systems and highlights areas where a data governance framework could make a meaningful impact on the state's ability to evaluate its programs. Through case studies from other states and specific recommendations for New Mexico, this report frames options for LESC to establish a greater level of data governance allowing the state's education data systems to better serve its long-term policy objectives.

## Hallmarks of High-Quality Data

The Legislature relies on data collected from New Mexico schools, often to make high stakes decisions about whether programs are serving their intended purpose. Relying on poor quality data to determine whether a program is effective can result in unintentionally funding ineffective programs, or perhaps worse, cutting programs that are truly effective. Experts agree data should meet certain standards to be considered "high-quality." For instance, tech industry leader [IBM](#) recommends organizations establish metrics in several domains to track the quality of institutional data, including the following:

- **Accuracy.** Data collected should accurately reflect their true nature in the real world. A student listed in a dataset should be as accurate a reflection of that student as possible, including the correct demographic characteristics, grade level, and outcomes.

### Key Takeaways

- Educational data often fails to meet the standards required to be considered "high-quality" (*Page 2*).
- Data quality issues have prevented LESC staff from conducting valid, reliable analyses of educational programs (*Pages 2-3*).
- Data governance, including internal PED policies as well as interagency data governance structures, can meaningfully reduce data quality issues (*Pages 4-5*).
- Other states, like California, Colorado, Kentucky, and Virginia, have established formal data governance structures to oversee their longitudinal data systems and provide strong, objective data analysis (*Pages 6-8*).
- LESC can pursue a range of data governance options during the 2025 legislative session, with varying degrees of complexity and effectiveness (*Page 8-9*).

- 
- **Completeness.** Data should contain all necessary records without missing values or gaps. This can be challenging when privacy laws like the federal Family Educational Rights and Privacy Act (FERPA) prevent the disclosure of data that could lead to individual students being identified.
  - **Consistency.** Data should be coherent and compatible across different datasets or systems. This can be as fundamental as ensuring that a dataset used for analysis includes the correct “ID” for the school participating in a particular program.
  - **Granularity.** The level of data collected for analysis should align with the intent of a program. For example, if a program is intended to operate at the school level, data should be collected and reported at the school level, rather than at the district level.
  - **Timeliness.** Data should be up-to-date and relevant for decision-making purposes. For example, the Legislature relies on up-to-date enrollment information to project funding needs for the upcoming school year.

While there are many other possible standards that could be used to measure data quality, these characteristics represent the minimum prerequisites data should meet to be considered high-quality. Unfortunately, due to a number of issues in the state’s data systems and school capacity to manage data systems, New Mexico struggles to meet each of these standards.

## Data for Diverse Purposes

Stakeholders across New Mexico’s education system rely on data to meet a variety of needs. According to the [Data Quality Campaign](#), families need data to answer big questions about students’ journeys through high school and into the work force, school and school district leaders need information that allows them to quickly target interventions where they are needed, and policymakers need data about whether their investments are having their intended effects. Indeed, even state education agencies need data to guide implementation; understanding how schools might be struggling to implement a program can help PED build effective, responsive guidance.

PED maintains a broad set of data systems and dashboards, but those systems are not designed to meet the needs of every stakeholder. For example, data collected in PED’s Nova system may be created to serve the department’s purposes of monitoring student enrollment, validating course load requirements, and ensuring compliance with state regulations. However, data collected in these systems may lack the reliability or the flexibility to answer broader questions posed by researchers, legislative analysts, or policymakers. In addition, datasets shared on the department’s website can be difficult to access, and even if these data are accessible, they may not be applicable to the contexts of the stakeholders that need the data. As researchers for the committee, LESC staff use data for specific purposes, and often encounter issues where messy data affects the outcomes of analysis.

## The Researcher’s Perspective: The Effects of Messy Data

At the beginning of the 2024 interim, LESC approved an ambitious [work plan](#) with the intent to study several state-funded initiatives, including the family income index (FII), structured literacy, career and technical education (CTE), and teacher clinical practice models. The goal of each of these evaluations was to learn more about the conditions under which the programs have been effective, and whether policy or budget should be adjusted to facilitate strong outcomes.

During each of these projects, issues related to data quality presented significant challenges. The analysis of FII schools was hampered by a lack of granular, relevant data that may have made it possible to see how differences in school spending patterns contributed to improved student outcomes. LESC’s analysis of structured literacy had to change course halfway through the evaluation due to data quality issues, shifting from a teacher-focused analysis of the impacts of structured literacy training to a less reliable schoolwide analysis of whether “model” and “support” schools meaningfully impacted overall student proficiency rates. LESC’s analysis of CTE programs lacked granular, relevant data on student participation in CTE programs, which is paramount to understanding how CTE affected students’ engagement, attendance, academic outcomes, and even holistic measures of success after students leave high school. In each of these studies, analysis of the available data resulted in a resounding answer of “it depends,” with some FII schools, structured literacy model schools, and CTE districts showing strong results, while others did not.

**Table 1: Examples of “Messy” and “Clean” Data**

Messy Dataset				Clean Dataset			
EPP ID	Name	Program	Graduation	State ID	Name	Program	Graduation
@1234	Benjamin Franklin	Traditional	A 2022	1001	Franklin, Benjamin	Traditional	Fall 2022
W239	Madison, James	BA-EE	Fall 2022	1002	Madison, James	Traditional	Fall 2022
XXXXX	Alexander Hamilton	GC	In progress	1003	Hamilton, Alexander	Traditional	NA
15093	GEORGE W	Alternative	Spring 21	1004	Washington, George	Alternative	Spring 2021
12334	John Q. Adams	Alt-Elem	Fall 2022	1005	Adams, John	Alternative	Fall 2022
90009	Samuel (Sam) Adams	Alternative	Fall 23	1006	Adams, Samuel	Alternative	Fall 2023

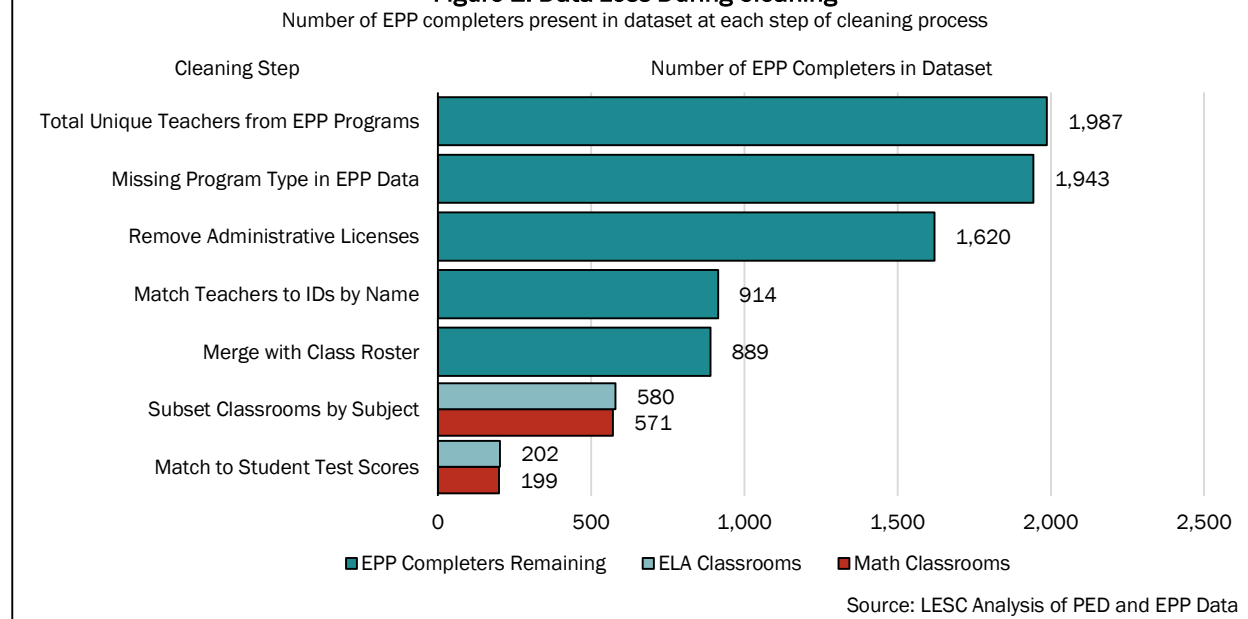
Source: LESC Files

LESC’s most recent report on teacher clinical practice was burdened by poor quality data from its outset. **Appendix A: Data Quality Issues Related to LESC Research on Teacher Clinical Practice** details how datasets provided both by PED and by New Mexico’s educator preparation programs (EPPs) violated the hallmarks of high-quality data introduced previously. The lack of clean, consistent data from EPPs and from PED also made the analysis of teacher residents’ outcomes nearly impossible.

LESC adopted a research question asking whether teachers in four distinct preparation pathways—traditional (student teaching), traditional (residency), alternative (teacher of record), and alternative (residency)—produced greater levels of student growth in one year of teaching. However, data available to PED on these four categories was messy and inaccurate, and in some cases, missing altogether. Finding PED’s dataset to be unusable, LESC staff instead relied on data provided directly from EPPs. However, these datasets were also messy, requiring a great deal of manual cleaning. **Table 1: Examples of “Messy” and “Clean” Data** provides examples of actual inconsistencies LESC staff found in EPP data with names anonymized. Inconsistent data reporting practices among the seven EPPs led to dozens of hours of manual transformation and several back-and-forth emails with EPP staff asking about the meaning of particular codes or whether names were complete.

As shown in **Figure 1: Data Loss During Cleaning**, each time LESC staff made a decision that resulted in cleaner, more reliable data, EPP completers were omitted from the analysis. For instance, after finding that teachers did not have a unique ID available to tie them from the EPP data to the state-level datasets, LESC staff determined the only way to tie EPP data to PED data was to use teachers’ first and last names. This is a problematic practice for a number of reasons, including the presence of duplicate names among teachers in New Mexico, the fact that teachers may get married or change names after they complete their EPP program, or the fact that the same person’s name may differ between datasets, like “James” and “Jim.” LESC staff used a fuzzy matching algorithm,

**Figure 1: Data Loss During Cleaning**



attempting to keep as many names as possible while preserving the fidelity of the data. Ultimately, while LESC staff expected to see more than 200 teacher residents in New Mexico classrooms by the 2023-2024 school year, staff were only able to analyze outcomes for just over 30 teacher residents, alongside about 170 completers that did not participate in a residency.

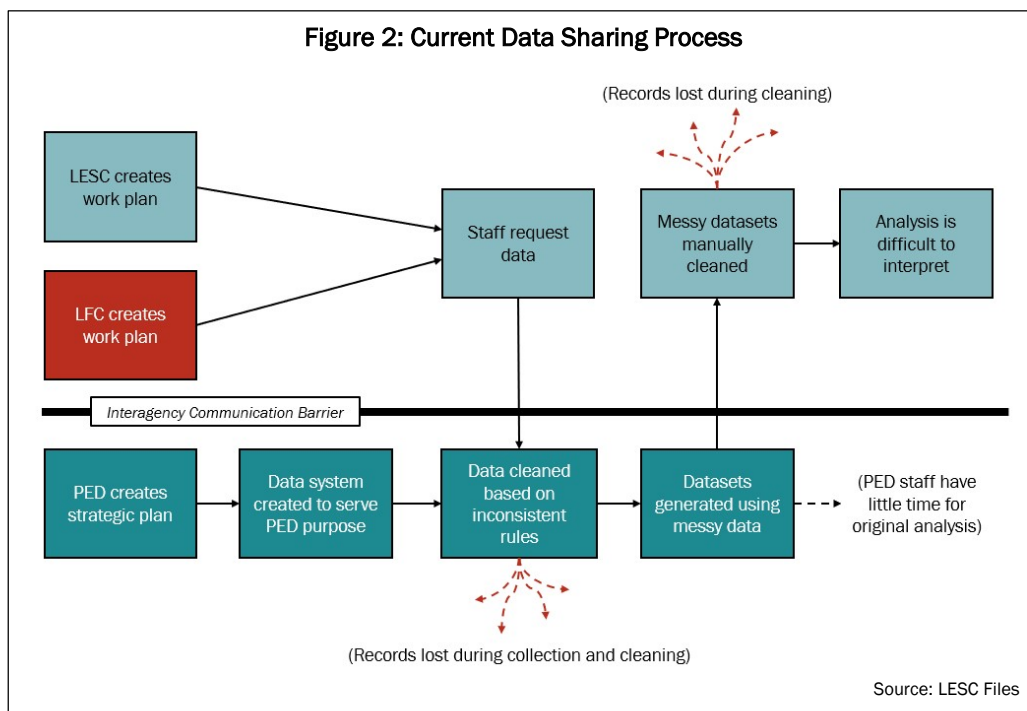
## Data Sharing Process

The existing process for collecting and sharing data with legislative staff, illustrated in **Figure 2: Current Data Sharing Process**, is ineffective. Several elements of the current system point to a greater need for collaboration and coordination among state agencies, including the following:

**Legislative committees create their research agendas in a vacuum.** Legislative committees have an interest in determining whether the programs they invest in are effective; such research is necessary to understand whether to scale programs up or cut them. Ideally, programs would be designed in a collaborative manner between LESC, the Legislative Finance Committee, PED, and other stakeholders responsible for implementing the programs. This type of collaboration could contribute to a common understanding of the goals of a program, and in turn, the data needed to evaluate whether those goals are being met. However, under the current process, LESC and LFC create their work plans in a vacuum, separate from PED’s work to create and adhere to its strategic plan. While there is some overlap in the three bodies strategic goals, the lack of unified state goals for the direction of education in New Mexico results in mixed priorities for which data should be collected and maintained.

**PED data systems primarily serve operational needs.** As the state education agency, PED primarily creates its data systems based on its internal operational needs. Data collected in Nova or in the Operating Budget Management System (OBMS) are primarily designed to track enrollment or monitor compliance, rather than to serve as tools to research how districts’ spending decisions drive student outcomes. Sometimes, this results in data being reported and used for purposes that were not originally intended.

**PED maintains many data systems across its many bureaus.** Separate bureaus in PED are responsible for the management of separate data systems. Nova is managed by the department’s information technology arm, OBMS is maintained by the school budget bureau, and school assessment results are managed by the Assessments, Research, Evaluation, and Accountability bureau. In addition, bureaus responsible for specific programs, like the FII or structured literacy, maintain their own spreadsheets regarding the program’s implementation. There is no set of overarching rules to ensure data are collected and cleaned in a similar way;



during each bureau’s disparate cleaning and validation process, records may be lost or information may be omitted (such as teacher ID numbers), impacting the quality and usability of the data.

**PED’s Policy Bureau is heavily burdened by legislative requests.** Given the lack of unified standards for data quality, PED staff spend a considerable time manually transforming data to fulfill legislative requests, often using messy data collected across several different data systems. As a result, PED staff often have little time to perform their own, original analysis of education data. Since June of 2024, LESC and LFC staff have submitted more than 120 requests for data or meetings with PED staff, which equates to about 6 requests per week. Many of legislative staffs’ requests are for the same datasets every year; an online portal for researchers to access common reports from project Nova would significantly reduce the number of requests from legislative staff.

To improve the interoperability of the state’s data systems, New Mexico should begin to consider a robust data governance framework. A strong data governance framework could ensure New Mexico’s data systems follow consistent standards and processes to support both PED’s operational needs as well as the state’s long-term strategic analysis. By prioritizing data governance, New Mexico’s education data systems can be better positioned to support informed decision-making across all levels of the education system.

## Data Governance to Improve Data Quality

Data governance is an umbrella term referring to an organization’s administrative oversight designed to make data more usable, accurate, and secure. The term “data governance” does not have a single definition, with many information technology organizations offering their own definition of how to make data governance effective. For instance, [Microsoft](#) defines data governance as “a system of internal policies that organizations use to manage, access, and secure enterprise data.” [Google](#) includes additional details in their definition, stating “data governance is everything you do to ensure data is secure, private, accurate, available, and usable, [including] the actions people must take, the processes they must follow, and the technology that supports them throughout the data life cycle.”

According to the National Center for Education Statistics (NCES), education data governance initiatives typically aim to improve data quality by focusing on three domains: greater organizational coordination, higher quality data, and improved usability of data. The benefits on each area of focus are listed in **Table 2: Key Benefits of Data Governance Initiatives**. In many cases, the benefits of data governance can be tied directly to the hallmarks of high quality data introduced early in this report, and also stand as a critical check to reduce the prevalence of data issues standing in the way of improved student outcomes.

**Table 2: Key Benefits of Data Governance Initiatives**

Greater Organizational Coordination	Higher Quality Data	Improved Usability of Data
<ul style="list-style-type: none"> <li>Establish clear ownership and responsibilities</li> <li>Reduce and eliminates redundant efforts</li> <li>Facilitate more frequent, better quality communication</li> <li>Standardize business practices over time</li> </ul>	<ul style="list-style-type: none"> <li>Bring errors and inconsistencies to light</li> <li>Improve accuracy and reliability of data</li> </ul>	<ul style="list-style-type: none"> <li>Provide timelier access to data</li> <li>Increase data security</li> <li>Improve how data is used to inform practice</li> </ul>

Source: NCES

As noted in Appendix A, a coordinated data governance structure along with a set of rules and policies governing data quality can improve data sharing practices in a tangible manner. Many of the issues that stood in the way of a high-quality evaluation of New Mexico’s teacher residency program could have been resolved given stronger coordination to eliminate inconsistencies.

- A state data governance oversight body could contribute to a [statewide research agenda](#), aligning state agency priorities for what data should be collected and used for evaluation.

- Data governance trainings can improve staff capacity to evaluate the accuracy of data.
- Data governance policies could require school districts, EPPs, or other reporting entities to report the same data to PED on an annual basis.
- Data governance policies could require datasets be examined for completeness before they are published.
- Data governance policies could ensure all reports include necessary information, like a unique ID for each unit of analysis.
- Data governance policies could require data to be shared in a timely manner, such as 30 days after the date of a request.
- Data governance policies could establish a more useful timeline for the validation of assessment data.
- Strong data governance will improve the quality of data, leading to more reliable analyses of programs and better investment of limited state funds.

## State AI Policy

During the 2024 interim, LESC staff studied the advent of artificial intelligence (AI) technology. As the technology grows and becomes more widespread, policymakers have an interest in ensuring the power introduced by AI is harnessed to improve the quality of education, rather than to act as a substitute for education.

Many state education agencies are relying on private and nonprofit guidance to establish policies governing the use of AI in education. For instance, TeachAI, an organization including representation from a number of technology and educational organizations, published a [Toolkit for AI in Education](#), a resource for teachers to better understand and utilize AI in classrooms, and for policymakers to craft AI policies.

To develop a strong, responsive state AI policy, national guidance should be blended with feedback from New Mexico communities. LESC staff participated in several AI policy engagement sessions with nonprofit Future Focused Education (FFE) during the 2024 legislative interim. Several themes emerged among participants regarding the use of AI in education. At a minimum, participants believed New Mexico's statewide AI policy should address the following topics:

- Consider the importance of a state-hosted AI system to keep New Mexico's data safe and tailor AI tools to New Mexico's cultural and educational contexts.
- Implement a framework that grants New Mexico's tribes, nations, pueblos, and other communities control over their data, including the option to remove data at any time.
- Ensure AI tools and benefits are accessible to all communities across New Mexico, with a focus on equitable representation.
- Ensure transparency in data collection and use.
- Consider whether AI should be provided as a public good, not controlled by private entities.
- Prevent AI tools from replacing meaningful human connection.

The Public Education Department has begun collecting feedback on the use of AI in education via stakeholder engagement sessions in three school districts this year. The department will use this feedback to help inform its guidance for schools.

Over time, it may be prudent to shift the responsibility of a state AI policy to a larger data governance board. AI is a relatively new technology, and as it advances and evolves, New Mexico's AI policy will need to be regularly assessed and updated incorporating feedback from both educators and technical experts. Legislation creating a statewide data governance board could include provisions requiring the board or an advisory subcommittee of the board to establish and maintain a statewide policy for the use of AI in education, and could include a set of minimum requirements such a policy should address.

---

## Data Governance Models in Other States

In some instances, legislation may be necessary to guide agencies' policies. This is particularly true for governance of data that spans across multiple state agencies. California, Colorado, Kentucky, and Virginia have all passed legislation prioritizing data governance, especially as it relates to interagency cooperation to create longitudinal prekindergarten through workforce data systems.

### California

In July 2021, California enacted [Assembly Bill 132](#), establishing the California Cradle-to-Career (C2C) Data System. The creation of the data system was the result of an 18-month planning process that consolidated feedback from more than 200 individuals across the California education system. During the planning process, stakeholders identified three shared purposes to inform how the system would be designed: tools for policymakers, researchers, educators, and advocates, tools for students and educators, and tools to support data use. Notable, the C2C system includes a query builder and research library accessible to the public, allowing researchers to quickly compile datasets of relevant information without the need for staff to manually respond to data requests. In addition, researchers can request access to restricted data and may be granted access if they meet certain confidentiality requirements.

The C2C data system is governed by a 21-member governance board, with membership enumerated in state statute. The governing board includes membership from the state superintendent of public instruction, institutions of higher education, the state's Student Aid Commission, the state's Commission on Teacher Credentialing, the Health and Human Services and Labor departments, educators, and legislators. The governing board is further subdivided into two advisory boards, one for "Data and Tools" tasked with examining whether the data system is providing actionable information and identifying ways to improve access to that information, and another for "Community Engagement," tasked with examining whether the Office of C2C Data is establishing strong feedback loops with end users and ensuring equitable access to information.

The governing board oversees the work of the "Office of Cradle-to-Career Data," an independent, neutral office created to initiate the data system, ensure "wide, appropriate, and legal use" of the data system to support policy researchers, scale tools to better serve educators, students, and families, and implement communications, professional development and technical assistance. The governing board is tasked with hiring an executive director for the office. Ultimately, the governing board is responsible for adopting a timeline for creation of the data system, ensuring the system is serving its intended purposes, and adopting data standards and security protocols.

To facilitate the creation of the data system and the work of the governing board, the California Assembly appropriated a total of \$2.5 million to several entities, including \$1.7 million to the state's department of education, as well as \$150 thousand to each of three institutions of higher education, the California student aid commission, and the state's Employment Development Department.

### Colorado

In 2024, Colorado enacted [House Bill 24-1364](#), creating a longitudinal data system and accompanying data governance board. The legislation explicitly states the purpose of the data system, grounding the measure in a recognized need for a highly educated, diverse workforce. The data system was the result of a 2021 task force that studied options to expand high school-to-career pathways and recommended measures to ensure equitable access to opportunities. The legislation mandates the system be created to facilitate a study of several Colorado workforce readiness programs.

To oversee the creation of the longitudinal data system, the bill established a governing board composed of five voting members and three advisory members. The voting members include the state Chief Information Officer, the executive director of the Department of Early Childhood, the Commissioner of Education, the executive director of the Department of Higher Education, and the executive director of the Department of Labor and Employment. Advisory members include appointments by the Governor, the attorney general, and the state board for community colleges and occupational education.

To inform the board's work, the bill also created two advisory boards: a "Systems Build and Implementation Interagency Advisory Group" and a "Sustainability Interagency Advisory Group." The two advisory groups include

---

appointments from the state agencies serving on the governing board designated for their expertise in data management and governance within their respective agencies.

To create the data system, Colorado appropriated \$5.4 million for FY25. The appropriation is intended to cover seven full-time positions at the Department of Education, with about \$1 million earmarked for implementation of the data system and the task force recommendations. The legislation does not create an independent agency to oversee the creation of the longitudinal data system, nor does it earmark funds for the meetings of the data system governing board or advisory boards.

## Kentucky

The [Kentucky Center for Statistics \(KYSTATS\)](#) is a semi-independent education data agency in Kentucky, established in 2012 as part of Kentucky's efforts to develop and maintain the Kentucky Longitudinal Data System (KLDS). KYSTATS collects, analyzes, and links data across multiple sectors—particularly in education and labor—to produce reports and insights that inform policies and programs. This aligns with Kentucky's legislative intent to create transparent, reliable data sources that benefit stakeholders, including policymakers, educators, and the general public.

The work of KYSTATS is overseen by an independent board, composed of the state's commissioners of public education, higher education, health and family services, and postsecondary education. The board is tasked with developing a detailed data access and use policy for access to the KLDS, establishing KYSTATS' research agenda, nominating an executive director of KYSTATS (for appointment by the governor), overseeing compliance with FERPA, ensuring KYSTATS reports are distributed appropriately, and providing general oversight over KYSTATS' operations. The board is given statutory authority to form subcommittees or advisory councils as needed to accomplish its purposes.

Initially, KYSTATS focused on creating integrated longitudinal data systems to support education and workforce planning. Over time, its role expanded to include broader analytic responsibilities, especially as data became crucial for assessing education outcomes, workforce readiness, and employment trends across the state.

## Virginia

In 2024, the Virginia state legislature created a working group to assess and plan improvements to the [Virginia Longitudinal Data System \(VLDS\)](#). The legislation tasks the working group to evaluate both the current and future needs of the VLDS and the Workforce Data Trust, specifically considering potential consolidation or improvements to databases, costs associated with hosting and maintaining the databases, and the governance structure for data sharing. The VLDS is responsive to a consolidated [statewide research agenda](#), which includes research questions like “what factors drive variation in the access to and availability of opportunities (e.g., housing, employment, education, ability and healthcare) in Virginia?” and “which programs or policies have been most successful at preparing students for higher education and for the workforce, and why?”

Unlike the other states noted above, Virginia has been implementing its VLDS since 2012. In the years after the VLDS was created, a number of disparate data governance structures were created across the many agencies that contribute to the VLDS. Over time, the lack of cohesion among the state's systems created challenges for agencies to contribute to the system. The fragmented approach to governance in Virginia—in some ways, an “over-governance” of the data system—resulted in less access to data. The Virginia task force was created to take stock of the existing structures for data governance, and to create a plan for a consolidated, streamlined governance structure responsive to the needs of the system.

By Virginia statute, the working group is responsible for creating a work plan that will include a timeline for database consolidation or enhancement, an assessment of necessary funding, staffing, and resources for maintenance, a governance structure for managing the databases, and designation of an entity to lead implementation. The work plan was scheduled to be released November 1, 2024.



---

## Policy Considerations and Recommendations

LESC is positioned to take action on education data governance during the 2025 legislative session. The Legislature's options lie along a continuum of complexity and effectiveness, with some of the more difficult governance policies perhaps reaching a greater level of effectiveness. The following considerations are ranked from low-complexity, with perhaps the lowest levels effectiveness, to high-complexity but with potentially the highest effectiveness.

- 1. Establish statutory standards and deadlines for data sharing.** A simple way to ensure data sharing occurs in a timely manner is to create a set of standards data must meet to be considered “ready to share” with legislative staff or other research institutions. These criteria could also set deadlines for when datasets should be made available. However, this model may not be as effective, as it does not ensure PED or other state agencies have the resources or other capacities necessary to meet the statutory requirements. Other solutions may be more effective because the focus on the internal and interagency communication necessary to build individuals' capacity to create and maintain datasets.
- 2. Require PED to create internal data governance policies.** PED would benefit from a set of internal policies to establish cleanliness and consistency across its many data systems. These data governance policies could establish timelines for when data are collected, cleaned, and publicized, and may include a set of criteria data should contain before publishing. Establishing a data governance policy within a state agency does not require legislation; in fact, many state agencies develop data governance policies as a set of internal business rules. PED has already begun work to refine its internal processes related to the statewide real-time data system, Nova.
- 3. Create an education data governance task force.** Before diving deep into the creation of a data governance board, it may be appropriate to create a task force of individuals across the education data ecosystem to evaluate the purposes and needs of an education data task force. A task force could evaluate the work that has already taken place to create a data governance structure, could build relationships between state agency data personnel, and could rally support around some future data governance structure, whether that is a data governance board or even an independent data analysis agency.
- 4. Create an interagency data governance board.** Making data interoperable between state agencies is a significant undertaking that requires a much greater level of data governance. The Legislature may wish to establish a data governance board designed to standardize processes, protocols, and shared frameworks for data use across multiple agencies, as well as to create powerful analytical tools to access data. Linking data between PED, the Higher Education Department, the Early Childhood Education and Care Department, the Children Youth and Families Department, the Department of Workforce Solutions, and the Division of Vocational Rehabilitation could give state agencies the means to gain a comprehensive view of the factors influencing student outcomes. Many other states accompany the creation of their longitudinal data system with a statutory data governance board or council, with membership from the many state agencies that contribute data to the new system.
- 5. Establish an independent agency responsible for staffing a data governance board.** Sometimes, an independent data agency that serves as staff to the state's data governance council can play a crucial role in promoting transparency, accountability, and cross-agency insights. An independent data analysis agency would be less bound to political forces, providing neutral, data-driven analyses to inform policy and program decisions. Offering a data agency full independence from individual agencies' priorities allows the agency to focus solely on rigorous analysis and the integrity of statewide data. This setup ensures that data use aligns with broader strategic goals rather than being influenced by the specific objectives of any one agency. An independent data agency has several consolidates data from multiple sources under strict privacy protocols, it enables more comprehensive, longitudinal analyses that can track individuals' progress from early education through to employment.

## APPENDIX A: Data Quality Issues Related to LESC Research on Teacher Clinical Practice

<b>Research Question</b>	<p><u>LESC Work Plan</u>: What are the effects of teacher preparation pathways and clinical experiences on student growth as measured by student proficiency growth on state assessments?</p>			<p>An administrative data governance council could set a <b>statewide research agenda</b>, creating alignment among state agencies on what data should be collected and used for evaluation.</p>
<b>Variables</b>	<p><b>Independent Variable:</b> EPP completer participation in various preparation programs</p>	<p><b>Mitigating Variable:</b> Student's placement in a EPP completer's classroom</p>	<p><b>Dependent Variable:</b> Student growth after one year with a teacher</p>	
<b>Data Needed</b>	<p>The number of recent EPP completers in one of four pathways:</p> <ul style="list-style-type: none"> <li>Traditional – Student Teaching</li> <li>Traditional – Residency</li> <li>Alternative – Teacher of Record</li> <li>Alternative – Residency</li> </ul>	<p>SY24 "Class Roster" dataset, showing the unique students placed with individual teachers in one school year</p>	<p>Two years of student-level assessment results (SY23 and SY24)</p>	
<b>Accuracy</b>	<p>Data obtained from EPPs did not match data provided by PED</p> <p>Staff decided to use EPP data given the inaccuracies present in PED data, but even EPP data contained inaccuracies.</p>			<p>Data governance policies and training can <b>improve staff capacity</b> to evaluate the accuracy of their data.</p>
<b>Completeness</b>	<p>Some EPPs did not collect data for all completers</p> <p>Some completers present in EPP datasets were missing from PED datasets, and vice versa</p>	<p>Initial class roster dataset was missing ~40 percent of teachers; this was later corrected</p>	<p>State law only requires students to be tested in grades 3-8 and 11.</p>	<p>Data governance policies could require all EPPs to <b>report the same data</b> to PED on an annual basis.</p>
<b>Consistency</b>	<p>EPPs do not collect data on completers in a uniform manner</p> <p>EPP completers were reported using inconsistent unique IDs</p> <p>PED dataset did not include EPPs unique IDs</p>	<p>Initial class roster dataset did not include teachers' unique IDs; this was later corrected</p>		<p>Data governance policies could require datasets be <b>examined for completeness</b> before they are published and shared.</p>
<b>Granularity</b>				<p>Data governance policies could require all reports to <b>include necessary information</b>, like a Unique ID for each teacher.</p>
<b>Timeliness</b>	<p>EPP completer data was requested from PED on 7/12/24</p> <p>PED delivered incomplete EPP data on 8/14/24</p> <p>Due to inaccurate PED data, LESC relied on datasets directly from EPPs, which in some cases took several attempts to get reliable data</p>	<p>Class roster dataset was requested 8/2/24</p> <p>Incomplete dataset was delivered 10/10/24, 69 days after request</p> <p>Complete dataset was delivered 11/1/24, 91 days after request</p>	<p>Assessment results were requested 8/2/2024</p> <p>Preliminary assessment results were delivered 9/9/2024, 38 days after request</p> <p>Final assessment results were delivered 11/6/24, 7 days before report was due for publication</p>	<p>Data governance policies could require data to be <b>shared in a timely manner</b>, such as 30 days after the date of a request.</p>
<b>Results</b>	<ul style="list-style-type: none"> <li>LESC staff faced difficulty matching teachers across multiple datasets without Unique IDs. During each step of the analysis, teachers had to be omitted due to the lack of a match.</li> <li>Staff were able to analyze results for 200 teachers of a possible 2,000 (~10 percent). Of more than 200 teacher residents expected to be in New Mexico classrooms by the 2023-2024 school year, analysis was only possible on about 30 residents in ELA and 30 residents in math (~15 percent).</li> <li>Based on the limited sample of teachers in the available data, there is some weak evidence that teachers who participated in an alternative residency program had a greater impact on student growth over one school year than teachers who did not participate.</li> </ul>			<p>Strong data governance will improve the quality of data, leading to <b>more reliable analyses of programs</b> and better investment of limited state funds.</p>