

The Future of Public Data Lies in the States, But It's Complicated

As the federal government withdraws from public data collection, states cannot step into the breach without addressing common challenges and disparities.

As a federal statistical agency, the National Center for Education Statistics (NCES) has long provided essential data for assessing education and workforce outcomes. But while the federal government dismantles NCES and related programs, a critical data gap is emerging as the country faces urgent postpandemic challenges like the sharp decline and slow recovery in math and reading scores. Statewide longitudinal data systems (SLDSs), used by over 40 states and Washington, DC, could help fill this gap but require attention and funding to overcome challenges and discrepancies among them.

These systems often include individual-level data on student enrollment, test scores, progress and completion, and more. Many integrate pre-K–12, postsecondary, and workforce data, allowing for the measurement of educational and employment outcomes and evaluations of youth programs. For example, Ohio, Kentucky, Tennessee, Virginia, and New Jersey produce the Multi-State Postsecondary Report, which provides postsecondary completion and employment outcomes by institution, major, and demographic. The data in this report have also been used to support evidence-based policymaking.

As the federal government withdraws from public data collection, data users and policymakers hope and expect states to compensate. But the 50 states, the District of Columbia, tribal lands, and territories have widely varying data infrastructures that reflect significant differences in staffing, technical capacity, data definitions, and the ability to navigate complex legal and regulatory environments. SLDSs differ in data they collect, state program data they link (if any), resources they provide for proper data use, and whether they share data with other state agencies. If left unsupported or unchecked, and in the absence of the centralization and comprehensiveness federal programs provide, states risk diverging in their data capacity as they do in other educational and quality of life outcomes. These differences in data handling, if unaddressed, could further exacerbate wide disparities in resources, opportunity, and prosperity across the country.

Although SLDSs are increasingly thought of as well positioned for rebuilding a national education data ecosystem, work is needed for them to step into this role. While the challenges are many, three rise to the fore: establishing common definitions and standards;

dealing with the thicket of laws and regulations governing privacy; and addressing chronic understaffing in state agencies. Through lessons drawn from my partnerships and collaborations with various SLDSs and other state engagements, I've gained some insight into what it may take to help build the resources, training, and support needed to expand access to education and workforce data. The onus for success cannot be entirely on states themselves: Policymakers, practitioners, researchers, and users all have parts to play in supporting state systems.

States face data challenges across multiple scales

Historically, the federal government has served as the unifying authority for data standards. The summary of a 2025 roundtable on reimagining NCES convened by the American Statistical Association (ASA) emphasized the importance of shared language and common data definitions, noting “wide variation in state definitions, standards, and priorities, making cross-state comparisons very difficult if not impractical without federal management and systematization.” Ensuring

As the federal role diminishes, the question of who should set, maintain, and lead shared definitions and standards becomes unavoidable—particularly when data must be shared and linked across state lines.

Beyond definitions and standards, states must also navigate the issue of myriad legal and regulatory complexities around privacy when managing sensitive or personally identifiable data. Education, health, and workforce data are governed by overlapping and sometimes conflicting federal and state regulations and frameworks such as the Family Educational Rights and Privacy Act (FERPA), Health Insurance Portability and Accountability Act standards, and the California Consumer Privacy Act (CCPA).

In Kentucky, for example, a mother's maiden name appears on a child's birth certificate even if she has legally changed her name. This practice complicates efforts to link parent and child records right from the outset. At the time of birth, a child's parent or place of residence may be the only available identifiers, since social security numbers (SSNs) are typically assigned weeks later.

In the absence of the centralization and comprehensiveness federal programs provide, states risk diverging in their data capacity as they do in other educational and quality of life outcomes.

common definitions, participants said, is an “inherently federal” function, even for such seemingly basic terms as *graduation rate*, *attendance*, or *middle school*.

Identical words and phrases can carry different meanings across states or even across agencies within the same state (for example, how student cohorts are defined). If states are expected to develop and rely more heavily on their own data infrastructure, they must address not only internal silos—such as those between education and social services in the state—but also cross-state data interoperability to realize the benefits of national data, including understanding student and workforce migration across state lines.

As with any effort that brings together diverse groups—in this case, multiple state entities for SLDSs—differences in language and terminology are inevitable. But the issues go even deeper; the ASA summary noted that “creating comparable national statistics requires data structure and formatting standards, including details like reference and submission timelines, which is currently accomplished by NCES in its collections from states.”

Similar inconsistencies arise when attempting to connect children's data to either biological parents' information. While SSNs are often critical for creating accurate, high-quality data linkages, many states, such as Kentucky, choose not to store or share them because of privacy and confidentiality concerns. As a result, agencies may rely on weaker identifiers, and these linkages can quickly break down, for instance, in cases where parents separate or divorce, birth parents are not legal guardians, or children enter foster care and move frequently.

Privacy concerns and consent limitations only compound these issues. In most states, for example, data collection on children younger than age 13 requires parent or guardian consent. For youth aged 13 to 17, both the youth and their parent or guardian must provide consent. Once individuals turn 18, they are considered legal adults and can provide consent independently.

If SLDSs aim to expand their data collection to follow individuals from pre-K through high school and into early adulthood, they will need to reevaluate consent procedures and adopt privacy protections that evolve with

participants' changing legal status and contact information. Tracking individuals aged 18 to 22 is especially challenging at the population level, as they are often highly mobile. They are more easily followed if they are enrolled in college, the military, or government programs, making linkages between SLDSs and other state data systems even more critical.

Failing to address these constraints and legal considerations early can leave states with data that are legally sound yet poorly suited for the policy and program decisions they are meant to inform. Or conversely, the situation can leave them with high-quality datasets that cannot be used as intended without violating laws.

In addition to these complexities, many states face a more practical limitation: people. By 2023, many private and public sectors had recovered from the jobs lost due to the pandemic, but state and local government employment remained below prepandemic levels. State agencies remain significantly understaffed, with existing personnel stretched across multiple responsibilities. Some states' data staffs are particularly small: New Mexico, for example, has just four people, while Kentucky has over 30 full-time staff supporting

aligns with recommendations from a recent National Academies of Sciences, Engineering, and Medicine report that emphasizes the importance of “a shared language reflecting the concepts of risk, harm, and usefulness. Shared language also enables quantification of these concepts, enabling them to be considered when managing trade-offs.”

This need applies even in well-resourced state entities such as California's Cradle-to-Career Data System and the Texas Education Research Center. In both cases, when partnering with these groups, we began with training sessions to establish a shared taxonomy before providing technical assistance on data-sharing efforts. These sessions also clarify data governance by identifying relevant decisionmakers and pathways for data products and other output approvals. Although this process does not always result in a clear leader, it can help move the needle when the process might otherwise stagnate.

For a more sustainable solution, progress often requires a coordinating leader, even temporarily, to align stakeholders and sustain efforts. If the federal government

Beyond definitions and standards, states must also navigate the issue of myriad legal and regulatory complexities around privacy when managing sensitive or personally identifiable data.

their SLDS. Asking state teams to take on new initiatives—even ones critical to their communities—often means pushing already overburdened staff beyond their limits. And asking states to coordinate data collection for sharing will further burden states with limited resources.

These constraints shape how public policy experts and other external partners in the data ecosystem can effectively support states. While states need assistance, support must address pressing real-time problems rather than hypothetical future needs. Any effort to strengthen state data infrastructure must demonstrate immediate, tangible value that justifies the limited time and bandwidth state staff function under.

Ensuring no state is left behind

In the absence of federal standards, one practical starting point would be establishing shared taxonomies and core concepts early in an initiative. In my team's partnerships with SLDSs and other government entities, we typically dedicate an early meeting as a training session to establish a common vocabulary across participants. This approach

does not fill this need, another entity must step forward to coordinate states and establish shared data standards. This might be through common data taxonomies or an agreed framework for safe data sharing. Identifying that entity and ensuring it can lead successfully may not be a simple matter.

Special attention must be paid to the early stages of an initiative—not only to establish shared taxonomies and core concepts, but also to consider how laws and regulations may affect data collection, governance, and privacy. As colleagues at the Urban Institute have emphasized, having such conversations at the outset of any initiative is especially important at the state level, where agencies must navigate federal, state, tribal, and local laws simultaneously. This often requires consultation with legal counsel across multiple entities, even within the same state. We have observed that legal teams from different agencies may disagree on how the same statute or regulation applies to a given dataset; awareness of these disagreements at the outset can help in building a strategy to address them in the most effective way.

In our partnership with the California Cradle-to-Career Data System, we addressed this challenge by engaging a consultant who specializes in education data privacy law. She led a dedicated training session on FERPA and possible interactions with California's regulations, including CCPA. This shared legal grounding is essential as our state partners determine how synthetic data should be treated in institutional review board reviews and other governance processes.

The emphasis on putting time and energy into the start of any state initiative extends even earlier to how initiatives are first conceived. A state's limited capacity often means the initiative must first solve an immediate problem for the state, such as reducing burden. For example, my first SLDS partnership was with the Nebraska Statewide Workforce & Educational Reporting System (NSWERS). Like many SLDSs, NSWERS holds invaluable administrative data, but privacy concerns require lengthy application processes and legal agreements for researchers and other state agencies seeking access. In many cases, applicants may discover that the available data do not meet their needs, resulting

has championed the work. As with NSWERS, the current approach relies on highly customized, over-engineered solutions for individual requests, which has proven to be difficult to sustain over time. The NMLDS team lead views our partnership as a way to build his team's technical capacity—along with other teams in the New Mexico Higher Education Department—while developing a new data product that could substantially reduce recurring manual data requests from legislative and other state stakeholders.

States also face additional challenges such as funding new initiatives and sustaining multiyear investments. Additionally, they strain to maintain infrastructure that evolves and dates rapidly while navigating both the promise and risks of artificial intelligence for data governance and privacy. Expecting states to step into the breach left by federal inaction, and to work together effectively, will require creative solutions to all these issues.

External bodies formed by states and subject matter experts like the State Higher Education Executive Officers Association and Council of Chief State School Officers,

Any effort to strengthen state data infrastructure must demonstrate immediate, tangible value that justifies the limited time and bandwidth state staff function under.

in wasted time and frustration on both sides. My team's partnership with NSWERS provided training, code, and other technical resources to help NSWERS develop a synthetic dataset—that is, a dataset designed to imitate confidential datasets while limiting information about individuals—reducing the burden of repeated requests while preserving privacy protections.

NSWERS is a well-resourced agency relative to most SLDSs. As of March 2026, NSWERS has 10 dedicated staff members and two postdoctoral researchers, along with contract support from my team to further their goals with synthetic data generation. Many states operate with far fewer resources and still choose to engage in new initiatives. In these contexts, progress often hinges on the presence of a champion—someone willing to own, advance, and sustain an effort despite competing demands.

The New Mexico Longitudinal Data System (NMLDS), for example, has just four dedicated staff members. Despite its limited capacity, our collaboration with NMLDS mirrors the early stages of the NSWERS partnership and exists largely because the team lead

along with other resources that could help address the challenges outlined here, have been established to identify needs and create guidelines and best practices. Yet nothing motivates action as strongly as a federal law, congressional mandate, or extra funding—mechanisms these entities now lack but that the federal government offered before, and still has at its disposal.

The federal government may be stepping away, but preventing deeper divides across states requires confronting these issues directly. Assuming states can simply pull themselves up by their bootstraps is no way to ensure that they, and the people who rely on their data, will not be left behind.

Claire McKay Bowen is a senior fellow at the Urban Institute who leads its Data Governance and Privacy body of work, advancing technical and policy solutions to responsibly represent people in data. She is also an American Statistical Association fellow, recognized for her contributions to statistical data privacy and professional leadership.