

Data Infrastructure Barriers

— Deliverable 6 —

for

Coleridge Initiative

Project No. ADC-STEM-24-N8

Project Agreement
Holder

Nathan Barrett, Ph.D.

Project Team Technical POC

Nathan Barrett, Ph.D.

1740 Broadway

New York, NY 10019

859-396-0987

nathan.barrett@coleridgeinitiative.org

Submitted: 5/23/2025

Table of Contents

Introduction.....	3
Legal & Policy Challenges	3
Governance Challenges	4
Data Standards & Quality Challenges	5
Technical Infrastructure Challenges	6
Capacity & Resources Challenges	7
Culture & Trust Challenges	8
Use Case Alignment Challenges.....	8
Sustainability Challenges.....	9

America’s DataHub Consortium (ADC), a public-private partnership, implements research opportunities that support the strategic objectives of the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF). These results document research funded through ADC and are being shared to inform interested parties of ongoing activities and to encourage further discussion. Any opinions, findings, conclusions, or recommendations expressed above do not necessarily reflect the views of NCSES or NSF. Please send questions to ncsesweb@nsf.gov. This product has been reviewed for unauthorized disclosure of confidential information under NCSES-DRN26-049.

Introduction

As the demand for evidence-based policymaking grows, so too does the need for robust data infrastructure capable of supporting integrated, cross-sectoral insights. States across the U.S. are increasingly recognizing the value of linking administrative data across education, workforce, health, and social services to inform decisions that improve outcomes for individuals and communities. These efforts are particularly crucial for understanding the full trajectories of students, jobseekers, and families—who often interact with multiple public systems over time.

Within individual states, establishing cross-agency data infrastructure presents a series of challenges that require careful navigation of legal constraints, governance hurdles, technical incompatibility, and cultural resistance. While these intra-state efforts are complex, they serve as the foundational stepping stones for a more ambitious goal: building interoperable data systems that span state borders. Cross-state data integration is becoming increasingly important in a mobile society where individuals live, learn, and work across jurisdictions, and where regional labor markets and educational pipelines are not confined to state boundaries.

However, the move from within-state to cross-state infrastructure multiplies both the opportunities and the challenges. Differences in law, data standards, technology platforms, and political will can make collaboration difficult to initiate and sustain. Yet the potential benefits—such as more complete longitudinal data, enhanced policy coordination, and better support for populations that cross state lines—are substantial and well worth the effort.

This document explores the core challenges associated with establishing both within-state and cross-state data infrastructure, drawing out comparisons and identifying strategies that have proven effective in overcoming barriers. Organized by key domains—legal and policy, governance, data standards and quality, and technical infrastructure—this brief aims to support practitioners, policymakers, and technical leaders in building durable, secure, and impactful data systems across jurisdictions.

Legal & Policy Challenges

Within a single state, legal and policy barriers primarily stem from the complexity of interpreting and applying federal laws such as FERPA (Family Educational Rights and Privacy Act), HIPAA (Health Insurance Portability and Accountability Act), and various state-specific data protection laws. Agencies within a state may operate under different understandings of these laws, leading to inconsistencies in how data is collected, stored, and shared. Some agencies may require parental or participant consent before data sharing, while others may use administrative exceptions, creating confusion or resistance.

Furthermore, there is often no overarching legal framework that compels interagency collaboration on data issues. As a result, each agency may develop its own legal rationale for data protection, making it difficult to align policies or processes. The absence of a shared legal architecture frequently leads to overly conservative interpretations of privacy protections, which can inhibit data use even for clearly permissible purposes.

Cross-state data sharing adds multiple layers of legal complexity. Not only do different states interpret federal laws like FERPA and HIPAA differently, but each state also operates under its own legal and regulatory framework. Some states have enacted more restrictive data privacy laws or executive orders governing the handling of personally identifiable information, while others have developed more permissive environments that encourage data integration for public benefit.

These differences necessitate the negotiation of highly customized data use agreements that often include elaborate provisions to address jurisdictional liability, third-party access, breach notification responsibilities, and allowable data uses. The time and cost required to achieve mutual legal satisfaction can slow cross-state collaborations considerably, and in some cases, prevent them entirely.

Despite these legal challenges, some multi-state initiatives have made significant progress by developing shared frameworks to streamline collaboration. One example is the Multi-State Postsecondary Feedback report, which sought to better connect postsecondary and workforce systems across state borders. However, differing interpretations of FERPA complicated efforts to link education data across state lines.¹ A successful strategy included developing a model data use agreement (DUA) that was reviewed and endorsed by the legal teams of each participating state. This template not only standardized legal language but also built trust among partners.

A broader strategy includes convening cross-state legal working groups to collaboratively interpret laws and develop shared compliance templates. Some states have also appointed dedicated data privacy officers or general counsels with expertise in intergovernmental data sharing, which has improved coordination and reduced delays.

Governance Challenges

Within individual states, governance challenges often arise from the lack of a centralized authority with the mandate to lead data integration efforts. Agencies may frequently operate in silos with competing interests and priorities, making coordination difficult. Without a formal governance structure, decisions about data standards, access, and use are made in an ad hoc manner, often depending on interpersonal relationships or political will.

The absence of binding mandates or interagency agreements leads to uneven engagement and frequent delays. Agencies may also differ in their capacity or willingness to engage, particularly when data integration efforts do not support their core functions. While strong political leadership can play a key role, progress can be vulnerable without legislative authority or institutional frameworks to sustain momentum through leadership transitions.

In a cross-state context, governance becomes even more complex. There is no shared authority or governance structure that spans state boundaries, requiring the creation of federated governance

¹ https://kystats.ky.gov/Reports/Tableau/2023_MSPSR

models or interstate compacts. These governance models must accommodate differing political contexts, administrative procedures, and stakeholder priorities across states.

The difficulty in achieving consensus across jurisdictions often leads to a lowest-common-denominator approach in governance structures, which can limit effectiveness. States may also be concerned about giving up autonomy or exposing politically sensitive data, further complicating governance negotiations.

Despite these challenges, several initiatives have shown that cross-state governance is possible with the right structures and trust-building efforts. The Multi-State Data Collaborative (MSDC) offers a model where participating states established a multi-state governance board with rotating leadership.² This board provides a platform for shared decision-making and helps to align priorities and timelines. Trust-building measures, such as shared leadership roles and transparency protocols, are crucial to the board's success.

Other successful strategies include establishing memoranda of understanding that define roles, responsibilities, and dispute resolution mechanisms. States have also created cross-agency advisory committees with both technical and policy representatives to guide decision-making and ensure accountability.

Data Standards & Quality Challenges

Within states, inconsistencies in data definitions, classification systems, and measurement units present major hurdles to data integration. For example, educational institutions and workforce agencies may record credentials or employment outcomes using different codes or levels of granularity. These mismatches lead to difficulties in aligning records for analytical purposes, often requiring complex and resource-intensive data cleaning and transformation efforts.

Data quality is also highly variable across agencies. Some systems are updated in real time and include validation checks, while others rely on manual entry or batch updates with less frequent quality controls. These disparities can compromise the reliability and comparability of integrated data products. In addition, missing data and inconsistent data entry practices can disproportionately affect vulnerable populations, undermining equity analyses and policy responsiveness.

Cross-state data harmonization efforts magnify these challenges significantly. States may use entirely different taxonomies or classification systems for sectors such as education, employment, health, and social services. For instance, the definition of a "high-quality job" may vary from state to state, affecting metrics used in workforce outcomes. Additionally, states may track different indicators altogether or apply distinct inclusion criteria when building their datasets.

The lack of standardized metadata documentation compounds the problem, as states may not provide clear guidance on variable definitions, data provenance, or known limitations. Without

² <https://www.naswa.org/partnerships/multi-state-data-collaboratives/about>

such documentation, interpreting the data becomes error-prone and limits meaningful analysis across jurisdictions.

One promising strategy to address lack of data standardization involves the adoption of the Common Education Data Standards (CEDS) and other national or sectoral frameworks.³ The Multi-State Postsecondary Feedback Report implemented a data mapping exercise where each state's unique fields were translated into a shared schema, enabling meaningful aggregation and comparison.⁴

States can also engage in joint data inventory exercises to surface definitional differences early in the project timeline. Establishing working groups focused on specific domains (e.g., post-secondary enrollment, wage outcomes) allows technical experts to co-develop harmonization protocols and validation processes that balance consistency with local relevance.

Technical Infrastructure Challenges

Many state agencies continue to rely on legacy IT systems that were developed independently and often decades ago. These systems typically lack modern features such as APIs, cloud integration, and real-time data processing. They may also depend on proprietary software or outdated programming languages, making interoperability difficult. As a result, integrating these systems requires considerable investment in data extraction, transformation, and secure transfer protocols.

Security and compliance are other key concerns. Varying levels of cybersecurity maturity across agencies can pose significant risks in an integrated environment. Without common identity management, encryption, or logging standards, ensuring secure access and auditability across systems becomes a major barrier.

In cross-state collaborations, technical challenges grow even more complex. States often operate under different IT infrastructures, vendor contracts, and interpretations of federal compliance requirements. Some states may be locked into long-term agreements with specific cloud providers, while others follow distinct compliance frameworks. These discrepancies can create fundamental incompatibilities that must be addressed through middleware, parallel environments, or secure data enclaves.

Moreover, cross-state data sharing often demands higher levels of reliability and performance due to increased volumes of data and more complex user access scenarios. Scalability and latency issues become significant, particularly when linking administrative data with real-time or near-real-time analytics.

³ <https://ceds.ed.gov/>

⁴ https://kystats.ky.gov/Reports/Tableau/2023_MSPSR

The Coleridge Initiative addresses infrastructure incompatibility through a federated architecture using secure cloud-based enclaves. Each state retains control over its data environment while the integration of data is supported through governance policies and the Five-Safes Framework.⁵

Other strategies include joint procurement agreements that align technical standards across states or investing in shared infrastructure such as regional data hubs. These approaches reduce duplication, facilitate compliance alignment, and promote long-term sustainability through economies of scale.

Capacity & Resources Challenges

Many state agencies lack sufficient technical staff, such as data engineers, privacy officers, and statisticians, to support integrated data systems. This shortage is exacerbated by competing internal priorities, as agencies may struggle to allocate time and resources to long-term infrastructure projects. Budgetary constraints and staff turnover hinder sustained progress further.

Training and professional development are also limited. Without a coordinated approach to data literacy and technical upskilling, many staff members lack the tools to fully engage in data integration efforts. These gaps make it difficult to maintain momentum on complex projects that require long-term commitment and specialized expertise.

Additionally, project management, change management, and legal expertise are often under-resourced, further complicating multi-agency collaborations. These limitations slow down implementation, reduce the quality of outputs, and create dependencies on short-term consultants rather than sustainable internal capacity.

Capacity gaps are even more evident when trying to align states with vastly different resource levels. Some states may have dedicated data teams and modern systems, while others rely on outdated technology and minimal staff. This imbalance can stall collaboration if one or more partners cannot keep pace with project demands or cannot afford to participate without external support.

Coordinating staffing models, allocating shared costs, and establishing clear expectations for each state's contribution become more difficult. Disparities in political will and institutional culture further complicate the alignment of human and financial resources. Moreover, institutional inertia and inconsistent access to federal or philanthropic resources across states create further gaps in participation.

One effective strategy is pooling federal funding or philanthropic support to establish shared technical teams that serve multiple states. For example, regional collaboratives have hired central technical assistance providers to supplement state capacity, allowing projects to move forward without each state needing full internal staffing. Building out data fellowship programs and peer learning networks can also strengthen internal capacity and help retain institutional knowledge.

⁵ <https://coleridgeinitiative.org/administrative-data-research-facility>

Other strategies include joint training programs across states⁶, investing in a shared curriculum on data ethics and governance, and developing career pathways for public-sector data talent. Embedding capacity-building into project timelines and budgets ensures that infrastructure growth is matched by the human capital required to maintain and scale its use.

Culture & Trust Challenges

Trust is a common barrier even within a state. Agencies may guard their data due to concerns about misuse, public scrutiny, or loss of control. Past experiences with failed data initiatives can breed skepticism, and without a compelling shared purpose, it's difficult to gain buy-in.

Data culture can vary widely across agencies, with some embracing transparency and collaboration, and others prioritizing compliance and risk avoidance. This cultural divide can impede the development of shared tools, protocols, and analytic approaches. In some cases, leadership turnover can reverse progress or shift priorities away from collaborative efforts.

Establishing a common language around data privacy, value, and impact is necessary to build a culture of collaboration. Without this cultural alignment, even well-designed technical systems may go unused or underutilized.

Cross-state projects face the same trust issues but amplified. States may view each other as competitors for funding, talent, or policy recognition. Historical mistrust, political polarization, or misaligned leadership agendas can make collaboration fragile or entirely unviable. Concerns about misuse of shared data or lack of control over data interpretation can stall or derail progress.

Trust also extends to the public. Building public support for cross-state data use requires transparency, stakeholder engagement, and assurances that data will be used ethically and securely. Disparate levels of openness across states can cause tension, especially if some states are more risk-averse or constrained by public scrutiny.

Early successes can help build momentum. For example, starting with a narrow, low-risk use case (like cross-border job placement tracking) and demonstrating clear mutual benefit can open the door for deeper collaboration. Facilitated convenings, neutral intermediaries, and data stewardship training can also reinforce a culture of responsible collaboration. Storytelling that emphasizes shared values and positive outcomes can help foster long-term trust among partners and with the public.

Joint data principles, cross-state governance charters, and mutual accountability mechanisms—such as peer review and transparent reporting—can institutionalize trust-building and help weather changes in leadership or policy context.

Use Case Alignment Challenges

Even within a state, agencies often have different missions and therefore prioritize different data use cases. One agency might focus on early childhood education, while another targets

⁶ <https://coleridgeinitiative.org/applied-data-analytics-training-programs>

workforce reentry. Without a shared use case, it's difficult to justify the effort and risk associated with data integration. Misalignment in objectives can lead to delays, underutilized infrastructure, or poor analytic return on investment.

Some agencies may also lack clarity on how data can support their mission, resulting in passive or reluctant participation. Others may face pressure to demonstrate short-term impact, discouraging investments in infrastructure that enables long-term, systemic insights.

When multiple states are involved, these misalignments multiply. For instance, one state may prioritize education outcomes while another focuses on public health or workforce mobility. Coordinating data infrastructure to serve multiple, diverse goals requires careful planning and compromise.

Differing policy timelines, evaluation metrics, and target populations can also create challenges in defining joint priorities or designing actionable analyses. Without clear incentives and aligned benefits, participation may wane. Inconsistent stakeholder engagement and lack of cross-sector leadership also inhibit convergence on high-value use cases.

Starting with cross-cutting issues—such as tracking students who move across state lines or workers employed in neighboring states—can reveal shared interests. These common problems help justify shared infrastructure and foster a sense of joint purpose. Co-developing use cases through inclusive planning processes can also ensure that resulting data products meet the needs of all partners.

Developing a portfolio of use cases—ranging from quick wins to deeper, multi-year initiatives—can help maintain interest and momentum while building trust and demonstrating impact. Engaging end users, including practitioners and communities, ensures relevance and promotes sustainability.

Sustainability Challenges

Many within-state data projects start with a grant or political momentum but taper off when funding ends or leadership changes. Sustainability requires long-term planning, institutionalization, and integration into agency operations. Without a clear business model and ongoing budget allocation, infrastructure may atrophy.

Succession planning, performance monitoring, and embedding data practices into agency workflows can help sustain efforts beyond individual champions or funding cycles. Institutional champions, legislative mandates, and cross-agency governance bodies can stabilize progress and signal long-term commitment.

These challenges are magnified across states. With no central funding or mandated authority, cross-state efforts often rely on temporary commitments. Differences in budget cycles, leadership terms, and agency mandates further destabilize sustainability. Inter-state competition for funding and policy visibility can also undermine collaboration.

Additional risk factors include inconsistent communications infrastructure, limited policy coherence, and overreliance on a small number of funders or anchor organizations. These issues can lead to fragmentation or collapse when partnerships face stress.

Institutionalizing the work through interstate compacts, nonprofit backbones, or national anchor institutions can improve sustainability. For example, placing the coordination function within a regional organization with long-term funding and neutrality has helped sustain several multi-state collaboratives over time.

Creating shared performance goals and embedding evaluation frameworks into partner agencies' strategic plans can also reinforce long-term commitment. Sustainability also depends on demonstrating ongoing value—through dashboards, reports, user stories, or other vehicles of dissemination—that justifies recurring investment. Multi-year roadmaps and shared funding mechanisms (e.g., pooled state contributions or federal matching funds) offer promising models for long-term viability.