COLORADO
Department of Higher Education

Colorado's Longitudinal Data Landscape

2023

Prepared and submitted by the
Colorado Department of Higher Education (CDHE)

Report to the Education Committees of the Colorado House of Representatives and the Colorado Senate
Statute: 23-1-141

January 2023

1600 Broadway, Suite 2200, Denver, CO 80202

For more information contact:
CDHE@dhe.state.co.us

Contents

Executive Summary

Colorado has led the nation in supporting the success of all Coloradans, enacting bold policies, and continuing to build a Colorado for all. These efforts have all been informed by robust stakeholder engagement paired with data analysis. Various initiatives over the years have improved the state's ability to use data to inform these goals. With support from historic investments in revitalizing the state's workforce, **Colorado has an opportunity to leverage data in well-governed, responsible ways to elevate evidence-based decision-making and support these goals to become reality.**

This report serves as a background on the Colorado longitudinal data landscape, with a special focus on Colorado's status and opportunities. The report will guide on-going conversations on data integration and Colorado's efforts toward a Longitudinal Data System (LDS). With a LDS, policymakers and other decisionmakers in other states have answered questions, such as:

- How many high school graduates go directly to jobs?
- What is the workforce supply for industry jobs?
- How can we better predict our economic indicators at the local level?
- What might be the impact of health care and service needs over time?

The report provides a brief description of Longitudinal Data Systems (LDS), including what they are, what data lies within them, how they can be used, and their benefits. The report also highlights efforts in other states where more coordinated and aligned systems have been successful. These states have found success by:

- Defining a unifying goal focused on a state's societal and economic needs by investing in career skill development with a focus on jobs critical to the state
- Vesting ownership in a goal to support the creation and enhancement of a state data system
- Identifying an office or team outside of existing state agencies to facilitate cross-departmental / agency work with a stakeholder group representing diverse perspectives
    - Team priorities related to an LDS include:
        - Meeting with key stakeholders to gather requirements for priority use cases for data system
        - Conducting a gap analysis to understand the difference between current data system capabilities and requirements identified
        - Specifying how information will be secure with clarity regarding governance, structure, privacy, and security protocols
        - Developing a team to build data system(s) through an iterative process to avoid linear "waterfall" steps
- Identifying and empowering expedient removal of bureaucratic hurdles to data access / data sharing - with an emphasis on connecting the supply and demand elements within a state

economy, such as employment and wage data to education data and job availability to skills and education pathways

- Leadership action or executive order to direct state agencies toward a call to action to develop or enhance longitudinal data system structures within existing resources and capacity

The report also includes recommendations related to greater democratization of data, following national best practices, and more alignment of Colorado's current landscape.

Introduction

The stories, perspectives, and experiences of all Coloradans enrich our state. Each voice helps the state better serve all Coloradans and develop policies and pathways to meet their needs. This is especially true within the education and workforce continuum. Colorado's education, training, and career spaces must continually adapt and respond to an individual's needs. To best adapt to these needs and support interventions that best serve Coloradans, evidence-based decision-making should be leveraged wherever possible. However, no meaningful change in education and workforce outcomes is possible without improved and publicly accessible data.

A Longitudinal Data System (LDS) incorporates longitudinal data from multiple datasets in the education and workforce space. Longitudinal data is collected over an extended period, usually multiple years, from the same subject. This allows for many programs and interventions to be evaluated for effectiveness. Much of the work done on earlier statewide longitudinal data systems has been spurred by federal grants to the Colorado Department of Education (CDE), meant to "develop or improve their data systems needed to effectively measure the success of educational programs."

Several pieces of legislation are central to the movement towards an LDS.

- **HB22-1330 (Higher Education Student Success)** implements student success strategies across the state and charges the Colorado Commission on Higher Education (CCHE) with reimagining the role of postsecondary institutions in a post-pandemic world to build economic resiliency and strengthen the state's workforce. The Student Success and Workforce Revitalization Task Force's final report provides multiple recommendations related to more robust, aligned data systems to facilitate postsecondary and workforce success. These efforts provide a longitudinal framework and move towards an LDS in Colorado.
- **HB22-1349 (Postsecondary Student Success Data System)** puts greater emphasis on tracking workforce and career outcomes of postsecondary education and pathways. It recognizes the need for a data system that collects and stores these data securely.
- **HB21-1111 (Consent Collection Personal Information)** directs an advisory group to study where personally identifiable information (PII) is stored by state agencies throughout Colorado, to study entities that have access to personally identifiable information stored by state agencies, and to determine the costs and processes necessary to centralize the storage and protection of personally identifiable information. This led to a data inventory by agency, which gives a better idea of where PII is stored and determines costs and processes to centralize its storage. The Office of Information Technology (OIT) final report summarizes this work.

- **HB22-1295 (Department Early Childhood And Universal Preschool Program)** states that Colorado Department of Early Childhood (CDEC) must work with other agencies to examine long-term outcomes of children that participate in CDEC programming, including universal preschool. These outcomes include academic outcomes such as third grade reading and graduation rates.

Taken together, these bills show an imperative to move toward more concerted and comprehensive data collection on education and workforce outcomes, analyze data with new and nuanced metrics of success, and secure that data in a way that protects the privacy of individuals. The status of these efforts, as well as the opportunities and challenges that exist in Colorado, will be explored through the rest of this report.

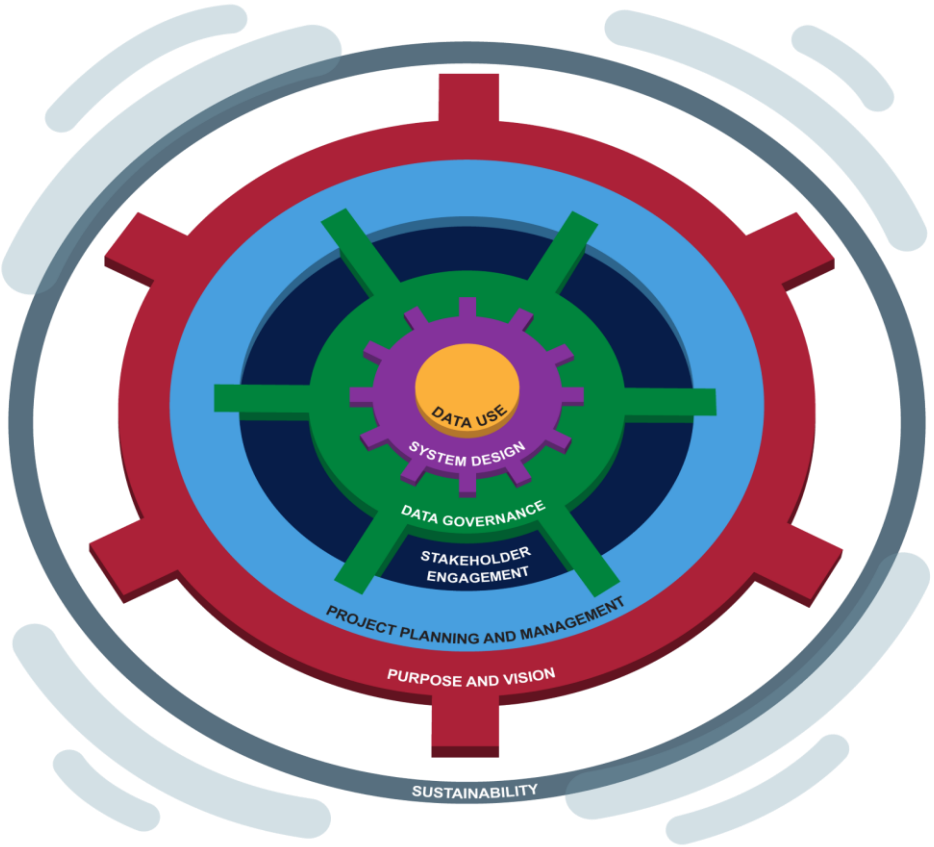
The report provides a brief description of Longitudinal Data Systems, including what they are, what data lies within them, how they can be used, and their benefits. Then, Colorado's efforts and initiatives related to these systems will be explored, particularly work done within the Colorado Data Trust, LINC, and the Colorado Department of Education. The report will summarize several challenges of LDSs including identity resolution, data harmonization, missing data, data usage/sharing/privacy concerns, culture changes required, and political and workforce challenges. Finally, the report will explore opportunities for growth and improvement in Colorado's LDS, highlighting other states' initiatives and best practices.

Longitudinal Data Systems

Overview and Framework

A Longitudinal Data System (LDS) incorporates longitudinal data from multiple datasets in the education and workforce space. Longitudinal data are data collected over an extended period, usually over multiple years. The data are collected from the same subject, allowing for many programs and interventions to be evaluated for effectiveness.

The Institute of Education Sciences (IES) established a state LDS (SLDS) framework to describe and ensure the essential components of an SLDS (Institute of Education Sciences). There are seven key components to this framework: 1) Purpose and Vision, 2) Project Planning and Management, 3) Stakeholder Engagement, 4) Data Governance, 5) System Design, 6) Data Use, and 7) Sustainability.



SLDS Framework (Institute of Education Sciences)

Purpose and vision consist of the reasons why the SLDS was built, the system's value, the scope of the SLDS, and how it is used. It can also include how the SLDS supports the state's long-term goals. This should inform all work that is done within the SLDS.

Project planning and management include detailed project plans, defined scopes of work, and documentation of roles, responsibilities, and procedures. This also entails the "execution of these plans, and adjustments to reflect changes in context, needs, and resources."

Stakeholder engagement looks to incorporate everyone contributing to or being served by the SLDS. All feedback must be valued and incorporated into plans, as it will "ensure the priorities, strategies, and execution respond to the expectations and needs of those it intends to serve."

Data governance reflects how "organizations or groups of organizations make decisions about their collective information assets," and is both an organizational process and structure. Strong data governance ensures that the decision-making process is representative of all entities within the SLDS, and "establishes responsibility for data quality, informs data use priorities, and provides a mechanism for establishing and continuously improving data policies and processes."

System design encompasses the technical components of the SLDS. The technology must satisfy the needs and functional requirements of the SLDS and align with the purpose and vision. This component will encounter ongoing maintenance throughout the life of the SLDS and may require modifications if needs or capabilities change.

Data use refers to people examining and making sense of the data to inform decisions and actions. This requires both prior knowledge and "experience with data to gain new knowledge that informs action." This can "inform stakeholders' work, measure program effectiveness, and shape state and local policies."

Sustainability reflects the efforts to "support and enhance an SLDS over time with sufficient human, organizational, and material resources to meet current and future needs." The three foundational aspects of sustainability in an SLDS are:

1) demonstration of value,
2) stakeholder support, and
3) capacity and resources.

Demonstration of value is straightforward; the SLDS must show that it serves a purpose and is useful to stakeholders. If not, stakeholder support, the second foundational aspect, will suffer. Lower support will further weaken the value of the SLDS. Lower support will also lead to less capacity and resources for the SLDS, which will of course lead to less value in the SLDS. These aspects can be thought of as intertwining and cyclical, as they all affect one another.

For our purposes, an LDS would include data from three core areas: K-12, postsecondary education, and workforce (Education Commission of the States). LDS would include data on the individual as well as the programs in which they are participating. Characteristics of the student or employee, such as race, sex, gender, ethnicity, ELL status, location within the state, and program participation, must be collected to both get a comprehensive picture of their experience, as well as analyze the equity or lack thereof for specific programs. We would also need education and enrollment data, specifically academic progress data like student achievement, test scores, growth over time, grades, classes taken, graduation status, postsecondary enrollment, and postsecondary credentials, as well as financial aid status including FAFSA (Free Application for Federal Student Aid), other federal aid, state aid, or specific program enrollment. This data would enable research on program status and effectiveness over time. We would also want workforce and employment data, including industry, occupation, alignment with credentials, and income. This would enable research on educational and occupational alignment, and the usefulness of specific credentials or training.

Federated vs. Centralized Approaches to LDS

There are two primary models for organizing an LDS: federated and centralized. A federated approach "temporarily links data from repositories maintained by participating organizations to create a report or to generate a dataset." The data is not located in a single repository, rather "individual source systems maintain control over their data but agree through MOUs or data sharing agreements to share the data with other participating systems upon request." This approach mirrors the systems currently in Colorado, specifically the Colorado Data Trust and CDE's Statewide Longitudinal Data System (SLDS), which are detailed later in the report. A federated model eases privacy concerns and allows data sharing agreements to be approved on time, but encounters some challenges in getting the data from agencies each time there is a research request, and greater difficulty when refreshing data (Institute of Education Sciences – Federated Data Systems).

A centralized approach is "a single, integrated data repository that contains, maintains, and provides secure access to data from all participating agencies and organizations." Here, the data are in a "single, centrally located data repository where they are organized, integrated, and stored using a common data standard." Kentucky is a notable example of this approach that we will detail further in the concluding section about other states' approaches. This approach allows data linkages from additional sources, but is reliant on data quality and timeliness from agencies and can suffer from lower trust amongst agencies given its centrality (Institute of Education Sciences – Centralized Data Systems). It should be noted that Kentucky uses a centralized model of K-12 IT management (the State Education Agency provides services to and exercises control over

districts around data and information management). This contrasts with Colorado's strong local control model, where school districts are independent in their data and IT decisions.

The chart below summarizes the pros and cons of both the federated and centralized model.

	Centralized Model	Federated Model
Pros	• Better performance for pulling data • More streamlined for data mining • Easier to account for data integrity/security • Single, central data policy • Easier to ensure data quality • Quicker data results • Avoids issues of disparate and noncompatible technologies	• Shorter time and lower cost for initial implementation • Mitigates turf battles and trust issues • Diffuses data and allows for tailored protection of data based on security • Quicker scalability
Cons	• Higher costs for infrastructure development and training • Data are only as current as the most recent load • Higher risk in the event of a breach due to the amount of data contained in a single repository • More difficult to distribute costs across participating agencies, if needed	• Requires data to be pulled and linked every time a dataset is generated, resulting in delayed results and potentially higher ongoing costs • More difficult to ensure consistent, quality results • Investment and support of intermediary interface by each of the participating agencies • Limited P-20W+ data integration

Table 2. Major pros and cons of centralized and federated data system models

Taken from SLDS Issue Brief: Centralized vs. Federated (IES)

Benefits of SLDS

There are multiple benefits to having a strong SLDS. The National Center for Education Statistics (NCES) points to four broad categories of benefits: 1) classroom and intervention insights, 2) administrative efficiency, 3) actionable research, and 4) workforce development (SLDS Grant Program Reasoning). First, a strong SLDS allows for greater insight into educational practices or interventions. With strong longitudinal data, we are empowered to get a more holistic view of programs, interventions, and policies that are meant to support learning. Instead of point-in-time insights that offer a view of how a program is functioning at one time, we can get longitudinal insights that show progression and development over time. Second, a strong SLDS can lead to greater administrative efficiency and "improve government efficiency by identifying redundancies and automating processes." This can include processes related to managing data requests or developing effective policies and processes. This lessens the amount of time for these requests to be carried out and increases the overall efficiency of how data collection and research flows. Third, a strong SLDS can lead to an increase in actionable research and program evaluation, which in turn can inform policymaking and resource allocation. Thus, research findings can be more accessible and usable, and collaboration with research groups can be increased. Finally, a strong SLDS can "promote workforce and economic development by informing the alignment of education pathways and employment opportunities." This includes addressing employment outcomes and workforce questions, achieving comparable employment outcome metrics, or modeling occupational pathways to inform state workforce supply and demand. This final benefit is, of course, dependent on each component of the framework being built and maintained thoughtfully and comprehensively.

Challenges with Longitudinal Data

Working with longitudinal data, and thus LDS, creates certain challenges. This data will usually require identity resolution, data restrictions, and political, workforce, and data infrastructure challenges.

Identity Resolution

Identity resolution is "a data management process through which an identity is searched and analyzed between disparate data sets and databases to find a match and/or resolve identities... [which] enables an organization to analyze a particular individual's or entity's identity based on its available data records and attributes." (Techopedia). A challenge arises when there is an inability to compare different identifier attributes, whether it be differences in the type of identifying attributes or the values that are provided. The data could have variations in spelling, formatting, or the type of attribute values referencing the same identity or lack consistent identifiers across systems, which then creates a chasm in linking different datasets. Changes in the individual or entity themselves can also create challenges; name changes or inconsistencies in abbreviations can make matching difficult as well. Thus, there is an inability to properly match records in some of these cases, as the identity of the record could be difficult to discern. If that is the case, there is difficulty in creating these larger longitudinal data sets, as the identity of some records can be difficult to match (Holland 2010).

The goal is to overcome spelling, phonetic, and other related errors and omissions in the data. One novel way to deal with identity resolution is through increased use of technology and matching scores to combine datasets. This can overcome some human errors or inconsistencies in identifier attributes. Data harmonization is also related to and required for identity resolution issues. Harmonization refers to transforming data from multiple forms, types, and sources, to create consistency in the data. This allows for comparability across individuals and creates a higher quality and more consistent data set (Rivery ELT Glossary). This includes having pre-defined field types to ensure consistency in data entry and standard procedures for data entry and sharing amongst agencies and divisions contributing to the LDS.

There is also a greater likelihood of missing data in these longitudinal data sets, especially if they rely on participant feedback or participation. This is not as common in administrative or state-collected data but can still occur if the data is collected from third-party collaborators or if tracking program or intervention effectiveness. When dealing with missing data, we want to glean insights from the dataset while not leading to less consistency or reliability. It is important to have standard procedures for missing data used consistently across datasets, as there will be

instances where the missing data will not be available to impute. Several strategies, including maximum likelihood estimation (MLE), multiple imputation (MI), or predictive mean matching (PMM) can work to address this missing data (Nooraee et al 2018).

Data Restrictions (Usage, Sharing, Privacy)

Perhaps one of the most crucial complications in LDS is the complexity of data usage and sharing agreements. Each data set or field can have different security and privacy requirements, and many data are contractually protected and must be used or shared only in particular ways. When developing an LDS, the data must also come from multiple agencies or divisions within the State. It is likely that each agency or division also has specific rules and/or laws about data usage and sharing, in addition to the dataset and field requirements.

A strong governance structure that utilizes data usage and use case agreements at an umbrella level is one way to remediate these challenges. With this structure, there may be no need for agency-agency data sharing agreements; instead, there can be a blanket agreement to join the LDS or Data Trust, with use-case agreements for specific data usage and projects. This is an example of a federated LDS, as "individual source systems maintain control over their data but agree through MOUs or data sharing agreements to share the data with other participating systems upon request." (IES Centralized vs. Federated). Data governance and sound privacy practices must be central to a strong LDS; data stewards must ensure that data is being used in ethically responsible ways that minimize privacy and sharing concerns. Data privacy and protection of an individual's data must be central to all efforts in maintaining well-governed, responsible data sharing.

The governance must also acknowledge the need for data privacy practices, which may include activities such as data destruction, data redisclosure, and how vendors and subcontractors are managed so that they also follow state and federal laws around data privacy.

There is also often a general lack of understanding of how data can be safely managed and protected. Explaining this complex environment can be challenging, so many people just assume it's too dangerous to take the risk of sharing data. Extra effort must be taken to ensure that stakeholders understand how their data will be managed and secured so that their privacy remains intact.

Political Challenges

Beyond challenges related to the data itself, efforts to create a Longitudinal Data System may encounter political challenges.

Clear public-facing information about the purpose of the LDS, how it will be used, and how individual information is protected is vital to building trust among all stakeholders for this work. Transparency must be built into all processes related to the LDS, including associated research projects and reports, and public input must be valued. The goal of the LDS should be to improve the connections between education and workforce outcomes and thus improve the opportunities that individuals in the state have. If individuals perceive the LDS as having different motivations, confidence will be eroded, and it will be more difficult to receive support. The best remedy for this is clear and concise information about the purposes, benefits, and safeguards of the LDS. It must be clear what the LDS is meant to do, how it will help others, and how individuals will be protected; otherwise, there will likely be little support for the LDS overall.

In addition, as mentioned above, the lack of clarity about data governance and proper decision-making processes for sharing data often leads to overly risk-averse approaches to sharing data hindering the effectiveness of the LDS. Proactive support from state and agency leadership indicating the importance of data sharing can help to overcome such obstacles.

Workforce Challenges

Related to the political challenges are workforce challenges. Without proper staffing to manage the LDS, and work to expand and improve the data, the LDS will suffer in quality and comprehensiveness. This will hinder the ability to gain valuable insights from the data, and thus the overall usefulness of the LDS.

The solutions are straightforward. First, there must be proper staffing, with positions dedicated to the management of the LDS. This has taken some form in Colorado through the creation and hiring of the Data Integration Manager position, which has taken the lead over the Colorado Data Trust. As the Colorado Data Trust continues to grow and incorporate more data sources, this position may need to be reorganized or extended. Another key solution is support from agency heads and leadership in the creation and improvement of the LDS. This should support further involvement in the Trust and more comprehensive research projects and contribute to the success and sustainability of the LDS. Each agency should also have a data steward that works with the LDS leadership on data quality and research; these stewards should aid in all research projects and work related to their agency, as well as help guide policies and procedures forward to ensure the sustainability and security of the LDS.

Data Infrastructure

Data Infrastructure relates closely to system design, as it refers to the technical components of the LDS. The technology must satisfy the needs and functional requirements of the LDS and align with the purpose and vision. It must also be able to implement privacy and security

requirements, data quality assurance, and validation. This is not to say that the data infrastructure should be solely reliable for these safeguards; rather, the infrastructure should be capable of flagging certain data-related issues and acknowledging privacy restrictions for these data. This is especially crucial when data is combined and integrated to create new datasets, as the original data sets may have different restrictions on them.

There must be a sustained investment in all aspects of the infrastructure to address these concerns, including both the technology and business processes. There are many potential providers of this infrastructure, all with different capabilities and costs. These solutions should be vetted to determine the best fit for the state. Once a technical infrastructure is in place, it must be evaluated consistently to establish whether all the state's needs are being met, as well as any areas that could benefit from modifications. The infrastructure will have to be maintained throughout its life, and much of this work will come from the vigilance of those working with the system, including LDS leadership, agency leadership, and data stewards. In addition to the technical information security infrastructure, security and privacy processes must be maintained to keep the LDS current and operational. The investment in sustainability is not just a technical one, the educational program side must also participate and stay up to date on data usage, management, and protection.

Colorado's Status

This section details the work that Colorado has undertaken over the years to make a Colorado LDS a reality.

Colorado Department of Education

The Colorado Department of Education (CDE) was awarded three Statewide Longitudinal Data Systems (SLDS) Grants. The first was in 2007 and the second was in 2009. The term SLDS is the official name of the US Department of Education's Preschool through Workforce grant program, in effect since 2005. The program was initiated by the National Center for Education Statistics (NCES) to help fund as many states as possible in their efforts to build technology systems to enable longitudinal data analysis to improve educational programs. Colorado has been an active participant in the national SLDS program since 2007 and was awarded a third grant in 2019. Each grant laid foundational elements toward the implementation of a true P-20W education data system that is aligned across the state's education systems and is anchored by a common definition of postsecondary and workforce readiness to ensure students graduating from high school are ready for postsecondary education as well as workforce success.

Through these grants, CDE developed data-linking technology and processes required to connect data between early childhood, K-12, higher ed, corrections, and the workforce. CDE partnered with the Colorado Departments of Higher Education (CDHE), Human Services (CDHS), Corrections (DOC), and Labor and Employment (CDLE) to test out use cases for each category of data listed above. As a result, Colorado successfully built a state longitudinal data system that meets the required system elements and capabilities as outlined in the SLDS grant program that funded the efforts. These initiatives, while built and tested as part of the 2009 SLDS grant, were not fully operationalized due to political and policy issues at the CDE and beyond.

CDE's technology matured in the years since the earlier SLDS grants, but the broader P20 data governance did not. The United States Department of Education (USDE) will be watching Colorado for improvements in cross-agency data governance with the most recent grant received in 2019. While CDE had the technology to share and use data for the benefit of education, its full potential was never realized. Work is currently underway to enable CDE to fully leverage the technology and longitudinally connected data to benefit education.

CDE learned many lessons from prior experiences in earlier SLDS grants and continues to be a partner in these collaborative efforts. In the meantime, CDE continues to move forward in its

efforts toward completing P20W statewide longitudinal data analysis with funding from the 2019 federal SLDS grant funding.

Colorado Data Trust

The Colorado Data Trust is a multi-agency, multi-partner agreement with three main goals:

1) Reducing barriers to ethical and responsible data sharing,
2) Encouraging collaboration and data usage to better serve the education and workforce pathways of Colorado residents, and
3) Moving towards more external data and resource sharing with both technical and non-technical stakeholders.

It is composed of a governance board, which is a decision-making body to ensure that trust agreements are being followed and is composed of at least one representative of each Data Trust member. The CWDC acts as the council and trustee, taking the role of convening and facilitating, as they supply no user data sets. While still evolving, there are defined processes to join and use the data, and current projects include My Colorado Journey, the CWDC Empowerment Score, Data for Opportunity in Occupation Reskilling Solutions (DOORS), Research Data Lake (RDL), and the Outcomes for Opportunity Workforce Dashboard. Agencies that wish to join the data trust must first sign the Colorado Data Trust Agreement. As a condition of their membership, the agency must identify one or more member-owned data resources from their division to contribute to the Data Trust, or at least make their metadata discoverable for all trust members. Once the agency is in the Data Trust, it can work to implement authorized uses for data within the Trust. Whoever wishes to implement a new authorized use case within the Trust must complete the Authorized Project Form in collaboration with Trust members and technical staff from the agency or agencies which own the requested data. This form will be shared with all voting members of the Colorado Data Trust, where the project will then either be approved or denied depending on the voting results. If approved, the project is added to the Data Trust Roadmap, and the sponsor of the project will work to provide regular updates at Trust meetings. The project will then be implemented with the respective data owners/stewards, vendors, and partner organizations to complete the project.

Within the Colorado Data Trust realm also exists a Research Data Lake created by Research Improving People's Lives (RIPL). This Research Data Lake (RDL) allows secure storage, organization, and combination of data sources across domains. It includes firewalls, encryption, and auditing tools to ensure security, and sensitive data is automatically anonymized with a global ID generated from PII at ingestion. This PII is then removed once the global ID is created, preserving confidentiality, and ensuring security. This data can then be used to glean insights and

information about programs or interventions in these individuals' lives and allows for the production of data dashboards or other public displays. The RDL rests on cloud technology and is highly flexible. Use of the data lake or multiple data lakes and additional relevant managed services could be expanded across the state to meet Colorado's LDS and other goals. The RDL is designed specifically to meet some of the challenges outlined above, including data quality assurance, data security and privacy, and sustainability. These cloud technologies are wholly owned and managed by state agencies, with support from RIPL as needed.

The data infrastructure is also improving, as the Trust is contracting with Brighthive on a new project management platform. This platform includes a data asset catalog that shows information about all the data in the Trust, member privileges that provide roles and permissions, and monitored and custom governance policies. The workspace will also allow filtering and sorting by project goals and success metrics and incorporates a project data map with capabilities to see what is happening with the data, who is providing assets, any issues with the pipeline or data linkages, and transformation steps including anonymization and aggregation.

PAIRIN's Records and Systems Integrator (RASI) is also a key feature in the Data Trust's landscape. It unifies learning, employment, and services records for career planning and case management and breaks down data silos by integrating data collected across a resident's lifetime (e.g., enrollment data, employment records, and social service records). This allows for tracking of longitudinal outcomes and combining de-identified, aggregated data based on any demographic or other information to track outcomes at scale.

The RASI is both a data lake and a collection of tools for facilitating communication among systems; matching, de-duplicating and cleansing participant records; and creating interactive real-time reports. The RASI secures sensitive data and provides access control to authorized users for case management and reporting. Any data the RASI needs is obtained "egress only," meaning the RASI can reach out into the internet and retrieve data to ingest, but external sources cannot directly access data within the RASI data lake.

Linked Information Network of Colorado (LINC)

LINC was established as a data collaborative in 2019 to securely connect datasets from multiple government partners and remove identifiable information so it can be used to make better decisions while protecting privacy. LINC is co-led by the Governor's Office of Information Technology and the Colorado Evaluation and Action Lab at the University of Denver. The LINC model is feasible through a three-tier legal framework that allows data partners to approve projects and contribute data without revisiting the same legal questions each time. This ensures data privacy and security are maintained while responding to data needs in a timely way. The

key to LINC's success for state and local agencies is a "flexible, respectful, and responsive" partnership model.

LINC works through the following four primary steps:

- **Project development**: LINC projects address real problems government partners face that demand analytics to inform smart decisions. Frequently these project ideas arise during government leadership meetings, task force convenings or informal brainstorming sessions. Researchers can also initiate LINC projects if they have expertise to contribute to an area of high priority to government partners.
- **Data partnership:** State government agencies benefit from LINC's streamlined, secure process to research complex policy, process, and service issues. LINC data partners can provide feedback and ultimately decide whether the project moves forward.
- **Link and de-identify data:** LINC is not a data warehouse. It is a federated data model where data partners temporarily provide datasets for an approved LINC project. The LINC Data Scientist secures the LINC project data, performs data linkages, and de-identifies the linked project data to meet the standards of all LINC data partners.
- **Ethical Data Usage:** The LINC Director provides continued oversight of LINC projects to ensure the LINC project data are only used for the approved purposes. The LINC data partners can review and share feedback before any results are shared publicly. LINC project data are destroyed once the project is over.

Since its launch, LINC has successfully produced actionable information for decision making on topics including youth homelessness, the early childhood workforce, perinatal substance exposure, and child welfare involvement prevention strategies. Because LINC is a low-cost, low-technology solution with quick turnaround, it can also be used as a first step in testing out the feasibility and value of more sophisticated data sharing solutions.

Assessment of Colorado's Efforts

This section describes the strengths and weaknesses of the work being done in the LDS landscape in Colorado. It begins with an assessment based on the SLDS framework described previously, then moves to more general paths forward.

SLDS Framework Assessment

To assess the efforts undertaken in Colorado thus far, it is important to use the framework we described on page 5. To review, the seven major components of this framework are: 1) purpose and vision, 2) project planning and management, 3) stakeholder engagement, 4) data governance, 5) system design, 6) data use, and 7) sustainability.

Colorado is mixed in its purpose and vision for the LDS landscape and the Colorado Data Trust. This vision has become more pertinent through efforts such as this report, as legislators see the value and purpose of an LDS overall. Both the Colorado Data Trust and LINC have defined mission statements that describe why the LDS is built, its values and scope, and how the LDS is used and supported. One area that can be improved is a better understanding of how the various players in this arena (the Colorado Data Trust and LINC) plan to work together. This would lead to a stronger purpose and vision for the landscape, rather than various pieces of the landscape.

Both the Colorado Data Trust and LINC have strong project planning and management structures, as described previously. These account for the data requested, how it will be used, who it will be used by, the tasks and activities involved with the project, and the overall execution of the project plan. There are procedures in place to ensure that projects are voted on and approved by these groups, as well as consistent updates and reflections while the project is ongoing. Again, one area that could be strengthened would be how LINC and the Data Trust should be working together. A clearer differentiation between the roles and responsibilities of LINC and the Data Trust would likely create more defined streams of work, and better collaboration opportunities between the two.
Stakeholder engagement could likely be improved. While agencies and nonprofits are engaged consistently, the public could be better informed about this work and why it is necessary. This could be through multiple means, but work is already happening to better engage this group in the Colorado Data Trust with public-facing documents and information shared online through the CWDC website. LINC offers great detail on its website into the purposes and procedures of its work. The research done by these groups should also result in public reports as often as possible (and appropriate). This will contribute to the transparency of these efforts, and could lead to

greater public buy-in. The Institute of Education Sciences offers a Stakeholder Engagement Plan Guide and Template that can help in these efforts moving forward.

Perhaps the greatest strength of the Colorado Data Trust and LINC is their commitment to strong data governance. This aligns with project planning and management. Both entities have consistent decision-making processes that include all affected parties, as well as guidelines on data quality, data use priorities, and general data policies and procedures. The Data Trust, for example, has set procedures for joining, implementing new authorized use cases, and using the data within the Trust. All projects must be voted on by members of the trust, and the Trust has the flexibility to include outside partners if sponsored by a governmental agency within the Trust. These governance procedures allow for flexibility while still ensuring sound data practices.

System design is mixed in Colorado; LINC has established a strong, low-tech infrastructure for their work, and the Colorado Data Trust is progressing to get to that point with the new project management software and back-end support that is currently making its way through contract negotiations. Once this platform is in place, the Colorado Data Trust will be more self-sufficient and less reliant on outside organizations for technical support. It will also allow for more consistent engagement from agencies on their specific projects and create a better technical solution that aligns with the purpose and vision of the work being done.

Data use is also mixed in Colorado, but has the potential to be transformative. Much of the data within the Colorado Data Trust and LINC is used to inform decisions and actions, whether it be the state or external partners. It does this through informing work, measuring program effectiveness, and shaping policies. There is still room to grow in this, both through more comprehensive data (from greater agency engagement) or increased frequency and topical breadth of reports. While this is continually a work in progress, this is something that the current landscape is doing moderately well given its resources and relative infancy. Several resources that could be helpful in this regard; the Data Use Standards Toolkit and Data Visualization Toolkit detail strategies and best practices related to data usage within an LDS.

The current landscape in Colorado is not sustainable without increased investment and clear commitment to this work. The landscape within the state cannot be conducive to enhancing an LDS over time without sustained human, organizational, and material resources committed to the work. The major weakness in Colorado's landscape currently is capacity and resources, as the Colorado Data Trust and LINC are not codified anywhere in statute and there has never been a significant, sustained, recurring state appropriation to do this work. The creation and hiring of the position of the Data Integration Manager serves this work somewhat, being a manager of the Data Trust and working on related initiatives, but this is a term-limited

position that isn't solely focused on LDS work. LINC collects project fees from the data recipients, but this presents challenges in responding to the needs of government partners that don't have budget lines that can cover the cost of high-priority LINC projects. If Colorado aims to create and sustain a single LDS, examples of other states should be followed. Many of these states have designated positions, or even offices, dedicated to this work and all that can come from it. These resources lead to greater capacity, and thus a demonstration of value and stakeholder support, which strengthen it further. The work done in Colorado has had a moderate demonstration of value through reports, and stakeholder support through external partners on work, but there is ample room to grow in each of these areas. More resources and a commitment to this work would help the LDS landscape in Colorado drastically. This evaluation is summarized in the table in Appendix A.

Other Suggestions

Colorado would benefit from more data and greater agency involvement in early education and the K-12 landscape. There is adequate representation in the higher education landscape (CDHE) and the workforce landscape (CWDC and CDLE), and the Colorado Department of Early Childhood is beginning to design and build an Early Childhood Integrated Data System, but the landscape is lacking otherwise. More agency engagement and data contributions would allow for more comprehensive research and insights.

There must also be a better understanding of how various initiatives work together, and a clear collaboration and data governance framework between these initiatives. The work being done between these initiatives should be complementary and build off one another. There should be a lack of redundancies and overlap of work, and a better framework for working together.

Best Practices for SLDS

Several best practices for LDS that could inform our efforts and work to improve the overall LDS landscape in Colorado are summarized by the Institute for Education Sciences (SLDS Best Practices Brief: P-20W+ Data Governance). These best practices can be separated into categories such as:

1) program sustainability,
2) data privacy,
3) engaged K-12 leaders and users, and
4) ongoing training and communication.

Program sustainability can be improved through training and support with the necessary infrastructure, including staffing for keeping it running in case of employment changes. It can

also be improved by not having it be at the whims of transitioning governors and/or legislators, as this political uncertainty can contribute to a lack of trust in the LDS and derail efforts to strengthen it.

Data privacy can be improved with strong protections and masking procedures, which are already in place in Colorado. One aspect that can contribute to greater transparency and privacy would be to grant access to people who have data in the system, likely through some public-facing dashboard. A great example of a public-facing dashboard that utilizes workforce insights is the Colorado Talent Dashboard. This dashboard differentiates between workforce supply and demand data, as well as access for legislators, program developers, job seekers and learners, and employers. This allows individuals to find the data that is most relevant to their aims, making it easy to use the dashboard and come back to it.

Engaged K-12 leaders and users can also improve an LDS. There must be leaders at all levels; this includes the state, county, school district, and school, and all leaders should be engaged in the work done with the LDS. Involving these individuals entails listening to their advice and building something that meets their needs. They must be kept engaged and involved throughout the life of the LDS.

Finally, ongoing training and communication are crucial for the LDS to operate as it's designed to. There must be training on data collection for those who are involved in the LDS daily. Further, it is important to coach end users, whether it be face-to-face or virtual, and gather feedback from teachers and administrators.

Other States' Initiatives

When we evaluate Colorado in these terms, it's clear that real progress has been made and continues to happen now. There are several places where the landscape in Colorado can be improved, namely system design, stakeholder engagement, and sustainability. By looking at other efforts for an LDS across the country, we can better understand the challenges and points of improvement that Colorado may face. Beyond the seven considerations of the LDS framework, it's important to focus on public access, the comprehensiveness of data, and transparency. We'll be focusing primarily on Connecticut, Hawaii, Kentucky, and Virginia.

Connecticut

Connecticut's efforts towards an SLDS are captured under their P20 Win Initiative within the Office of Policy and Management. As they state on their website, "P20 WIN informs sound policies and practice through the secure sharing of longitudinal data across the participating agencies to ensure that individuals successfully navigate supportive services and educational pathways in the workforce." Connecticut's system is an example of a federated data system, and has ten state agencies participating: "the Office of Early Childhood (OEC), the State Department of Education (SDE), Connecticut State Colleges and Universities (CSCU), the University of Connecticut (UCONN), the CT Conference of Independent Colleges (CCIC), the Department of Labor (DOL), the Department of Social Services (DSS), the Department of Children and Families (DCF), the Office of Higher Education (OHE), and the CT Coalition to End Homelessness (CCEH)." Each agency appoints a data steward to handle technical requests and work within the system.

Like the Colorado Data Trust and CDE's SLDS, P20 WIN has extensive data governance policies. There is an Executive Board and Data Governing Board "that develop and implement the necessary policies and procedures for multi-agency Data Sharing system to address broad policy questions and state needs." The Office of Policy and Management provides program management to support the operation and improvement of this initiative, and the Department of Labor conducts all data matching for approved requests.

P20 WIN also has a learning agenda that reflects priority issues for the State of Connecticut. There are five main avenues here: 1) College and Career Readiness, 2) Student Readiness, 3) Financial Aid, 4) Workforce training, and 5) Overcoming Barriers to Success.

All legal and technical agreements are hosted on the P20 WIN website, including past agreements that are no longer in effect. There are also several reports hosted on the website where individuals can read about the work that is being done and how it relates to the learning

agenda. Furthermore, there is a section to request data that outlines the process for starting a project within the P20 WIN initiative.

Together, Connecticut shows the potential for a federated SLDS. The comprehensiveness and transparency of the data and how it's used foster greater trust in the system and the work that's being done. This is paired with an ample amount of public information and access to all guiding documents, which further strengthens stakeholder engagement and data usage. This all contributes to the sustainability of the SLDS and sets Connecticut apart from many other efforts in the country.

Hawaii

Hawaii's efforts are also noteworthy. The Hawaii Data eXchange Partnership (DXP) is a partnership of five state agencies "committed to cross-agency data sharing to improve education and workforce outcomes in the state." The five agencies are 1) the Hawaii State Department of Education, 2) the University of Hawaii, 3) the Hawaii Department of Labor and Industrial Relations, 4) the Hawaii State Department of Health, and 5) the Hawaii Department of Human Services.

Hawaii describes its efforts through six main streams:

1) management of the SLDS,
2) shared data governance,
3) analysis and reports,
4) data products including dashboards, data stories, and publications,
5) data requests and provision to individuals and groups, and
6) a yearly data summit that convenes DXP stakeholders.

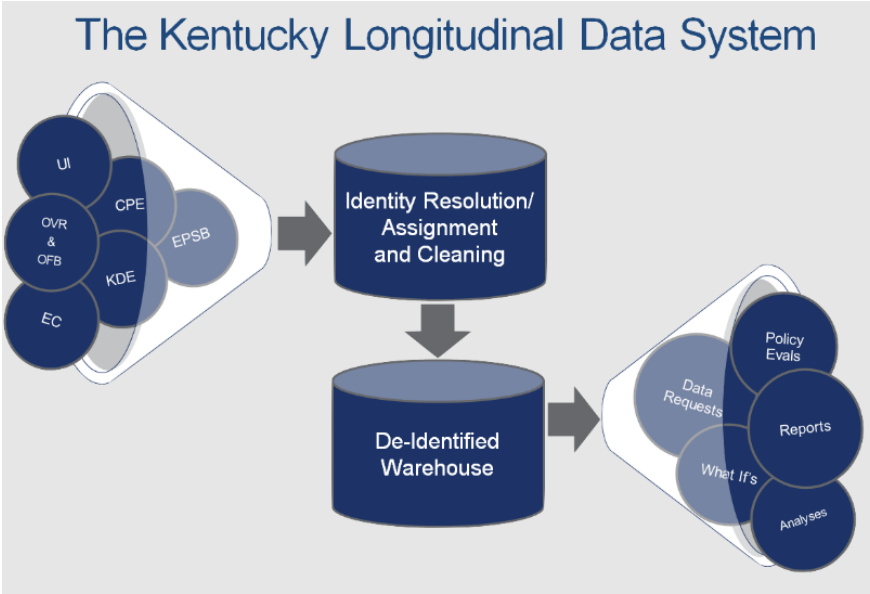
These relate to their three major goals within the state: 1) create a culture of data use, 2) strengthen the data sharing partnership, and 3) continue to protect individual privacy.

Hawaii stands out for its commitment to public access and stakeholder support, especially through its data and publications tab. Individuals can view dashboards, including the College and Career Readiness Indicators or the University of Hawaii Graduates in the Workforce, data stories including the growth of Dual Credit Opportunities in Hawaii, presentations, issue briefs that spotlight critical issues, or reports that more heavily explore topics within this landscape. You can also sort or filter by major area; you can view the data and any associated work that has been done by the early childhood, K-12, postsecondary, or workforce sector. Of course, there is plenty of information about data usage and requests, both at the aggregate and individual levels. There is also ample information on updates and upcoming events sponsored by the DXP.

The Hawaii Data Exchange Partnership (DXP) is managed by Hawaii P-20, which is an administrative unit under the Office of the Executive Vice President for Academic Affairs. There is a position dedicated to the project director of this work, as well as three institutional analysts, one data governance coordinator, one IT systems architect, and one IT specialist. There is an executive committee, a data governance and access committee, a research and data request sub-committee, and a security and access sub-committee. The responsibilities and roles of these committees are described here. Their data governance policy is described in detail here.

Kentucky

Kentucky is one of the gold standards of SLDS work, specifically a centralized approach. Known as the Kentucky Center for Statistics (KYSTATS), Kentucky works to collect and link data to evaluate education and workforce efforts within the state. This work includes the Kentucky Longitudinal Data System (KLDS), which is a statewide education and workforce data system. This work centers security and privacy at the forefront, which is detailed in the diagram below.



Taken from Security and Privacy - KYSTATS

Data is contributed by the Kentucky Department of Education (KDE), the Council on Postsecondary Education (CPE), the Education Professional Standards Board (EPSB), and the Kentucky Unemployment Insurance Claims System. KYSTATS has an executive director, four dedicated managers on the team, and 41 staff members dedicated to doing this work. There is a five-person board that decides on the overall direction of this work, and funding is provided through state appropriations, federal grants, user fees, and other grants or contributions from

public agencies or other entities. Funding from these sources has ranged over the years from approximately $3 million to $6 million.

Kentucky is also strong in data usage and stakeholder support. Their website includes a career explorer and family resource simulator that allows for more personalization using the data and work done in the SLDS. This shows the real value of this work to an individual that may be somewhat removed from the work and benefits from sustainability in the long-term.

Virginia

The Virginia Longitudinal Data System (VLDS) is a "P20+ system comprised of a core of data that is centered on education and workforce data from K-12, postsecondary, selected social services programs and health professional survey data." The VLDS includes the Virginia Department of Education (VDOE), the State Council of Higher Education for Virginia (SCHEV), the Virginia Employment Commission (VEC), the Virginia Department of Social Services (VDSS), the Virginia Community College System (VCCS), the Virginia Department for Aging and Rehabilitative Services (DARS), the Virginia Department of Juvenile Justice (DJJ), the Virginia Goodwill Network (VGN), the Virginia Department for the Blind and Vision Impaired (DBVI) and Virginia Department of Health Professions (DHP). It is an example of a federated model, and merges data in a "complex double-deidentifying hashing process that leaves private data behind the existing firewalls of the participating agencies (VLDS)."

The data governance of this work is incredibly strong as well. The VLDS Book of Data Governance outlines these efforts in large. Several committees lead work on this: 1) the Data Governance Council, 2) the Technical Subcommittee, 3) the Research Subcommittee, 4) the Communications Subcommittee, and 5) the Legislative Subcommittee. All the duties are detailed in the VLDS Book of Data Governance. There are also set procedures to become a VLDS partnering agency, including onboarding. The research process is also defined strictly, and consists of 1) receiving preliminary approval, 2) setting up a research proposal and gaining formal approval, 3) developing a research query, 4) conducting research with received data, and 5) communicating findings. This is similar to the processes set up in Colorado.

Stakeholder engagement and data usage are extraordinarily strong in Virginia. All the research that has been done within the VLDS is shared, as well as VLDS research that is in progress currently. There is also a great deal of information on privacy and how the VLDS addresses those concerns, which can help address the lack of trust in the system by those who live in the state. Again, this work contributes to a greater sense of sustainability within the SLDS, which cycles into greater stakeholder support and data usage. Multi-year grants have provided funding for Virginia's work over the years ranging from approximately $3 million to $17 million.

Moving Colorado Forward

Efforts to connect and align Colorado's longitudinal data landscape can yield important insights into better serving the needs of all Coloradans and allowing the stories of more Coloradans to be elevated. Additionally, the well-governed, responsible connection of Colorado data driven by robust questions show the value and importance of this work.

Democratizing Data

A central recommendation of this report is greater public access to the LDS system and insights gleaned from that system. There needs to be better differentiation of access that is specific to a user's needs. For instance, individuals should be able to access dashboards that allow them to view aggregate and trend information side-by-side with their data. This dashboard can enable students to navigate transitions from K-12 to postsecondary to workforce and career development resources, and allow job seekers to navigate career pathways that match their preferences and skills.

Researchers should be able to view dashboards, reports, and tools that display the federal and state government reporting indicators and other indicators that can be disaggregated by population and geographic region. This can enable prompt, efficient research and insights into the landscape of Colorado's talent development efforts.

Policymakers should have the functionality to allow them to investigate policy questions to direct workforce and education investments and programs. This will ensure that Colorado is staying on the path and investing resources where they are needed. There should also be linkages to other systems if considered helpful. One example of this is the Georgia Information Tunnel; this links the LDS to district-level student information systems (SIS) and "allows district administrators, principals, teachers, and parents to access state education data through their district's existing program" (Data Quality Campaign - Georgia Information Tunnels). Linkages to educational institutions or workforce boards can create greater usage of the LDS and avoid unnecessary duplication and investment in system building.

Following National Best Practices

There are several actions that Colorado can take to strengthen its LDS landscape. The Data Quality Campaign recently released a report on LDS that includes several recommendations to move LDS work forward (Data Quality Campaign 2022). The report outlines what is needed for this type of LDS access.

Political will is paramount to strengthening Colorado's LDS landscape, as "the ultimate success and sustainability of any user-centered LDS effort depend on intentional system design built from broad community engagement and trust, bolstered by executive-level leadership." This will lead to greater access to resources and greater sustainability for the LDS system. This could include codifying LDS work in statute. Coordinated leadership on this topic will drive greater collaboration among all efforts underway.

Governance, which was detailed earlier, is also necessary. Mirroring efforts by the Colorado Data Trust and LINC, there must be a state-level governing body composed of leadership from state agencies that supply data to the system and perform prioritization of data sharing activities. Taking lessons learned from CDE's SLDS progress, and similar to the Colorado Data Trust, it is recommended to have a group define each party's purposes, roles, and responsibilities and create a data policy that centers on privacy, equity, and timely access. Colorado would benefit from a greater understanding of how disparate initiatives that contribute to the LDS landscape work together and centering data governance in these discussions.

Privacy must also be central to any LDS effort. There should be clear rules and guidelines at the state level that center transparency for data users about what data is accessible under what circumstances. Privacy must be central to all data governance policies and priorities.

Technology infrastructure should allow users to securely access data without creating disparate, expensive tools across the state system. Ideally, the infrastructure would connect to other systems and should be implemented at scale. Colorado has made progress on this recently, contracting for an infrastructure platform that allows for greater coordination amongst agencies within the Data Trust. That said, more work must be done to allow individuals outside the trust to access this information. This can include a greater frequency of reports or real-time dashboards, as detailed previously. Hawaii is a notable example of centering public access to the work done by the LDS. Building upon current efforts within Colorado presents an opportunity for enhanced alignment.

There must also be a greater commitment from Colorado to **talent and human capacity**, and investment in the LDS landscape. There must be someone dedicated to leading the LDS work, beyond a limited-term employment assignment. There must also be an LDS governing body with a diverse staff to guide the work. Kentucky is a fitting example of this, having an agency solely dedicated to this work. Colorado must move towards this with greater **investment**, including "sustained funding such as line items in the state budget, dedicated grant programs, or blending and braiding of federal funding streams." This funding should support technological infrastructure along with human capacity, training, and development of the system.

Aligning Colorado's Landscape

As outlined in this report, a variety of efforts related to longitudinal data have emerged in Colorado over the years for various grants, projects, and uses. CDE's SLDS work, the Data Trust, and LINC are some examples and these efforts have accomplished important work that has provided important insights for all Coloradans. The Colorado Department of Early Childhood's ECIDS (Early Childhood Integrated Data System) is another tool in the overall Colorado state government longitudinal data projects belt, and a system that will continue to need to be fleshed out and improved to continue cross-agency collaboration. But when separated, the impact of these initiatives can be limited.

CDE's SLDS work has developed data-linking technology to accomplish projects. However, some of that work has been limited by a need for enhanced data governance processes and policies to facilitate better cross-agency collaboration. LINC has developed connections across a variety areas and processes for data-sharing and connection. However, LINC uses very manual data connection procedures for projects (by design due to some limitations from partners). The Data Trust and the related vendor partners have worked to leverage more advanced ways of connecting data for a variety of use cases. However, constrains around staff bandwidth and sustainable funding have limited the initiative's ability to execute on projects. Building off the strengths of each initiative in tandem with staffing and financial support (as summarized above in the Kentucky and Virginia examples) can allow the initiatives to have an impact greater than the sum of their parts.

By better connecting and aligning these efforts in more formal ways, a more coordinated ecosystem on longitudinal data can unlock insights to help policymakers better serve Coloradans and equip all Coloradans with more valuable information to thrive in their communities.

Appendix A: LDS Framework Evaluation for Colorado

	Strength	Points of Strength	Improvements
Purpose + Vision	Mixed	- legislative imperative - mission statements	- better understanding of how various initiatives work together (ex: CO Data Trust, LINC, and CDE's SLDS)
Project Planning + Management	Strong	- defined processes for project planning and statements of work - documentation and execution of roles, responsibilities, and procedures	- better understanding of how various initiatives work together (ex: CO Data Trust, LINC, and CDE's SLDS)
Stakeholder Engagement	Mixed	- incorporation of agencies participating in the LDS - feedback mechanisms of those participating in the LDS	- better incorporation of those being served by the LDS - more development of public-facing materials and mechanisms for influence - more engagement of research through public reports or memos
Data Governance	Strong	- decision-making processes representative of agencies - data quality and data use priorities are clearly defined - strong mechanisms for establishing and improving data policies and processes	- better understanding of how various initiatives work together (ex: CO Data Trust, LINC, and CDE's SLDS)
System Design	Somewhat Strong	- technical components satisfy the needs and requirements of the LDS	- reliant on an outside vendor for changes to infrastructure or contracting

		- a mechanism for improvements and maintenance	
Data Use	Mixed	- projects informing stakeholder work, measuring program effectiveness, and shaping policies	- increased frequency and topical breadth of reporting - more comprehensive data from increased agency involvement
Sustainability	Somewhat Weak	- limited demonstration of value through projects - improved commitment level by legislators as of late	- the need for recurring and sustained investment - designated position for the management of the LDS, beyond LTE - better understanding of how various initiatives work together (ex: CO Data Trust and LINC)