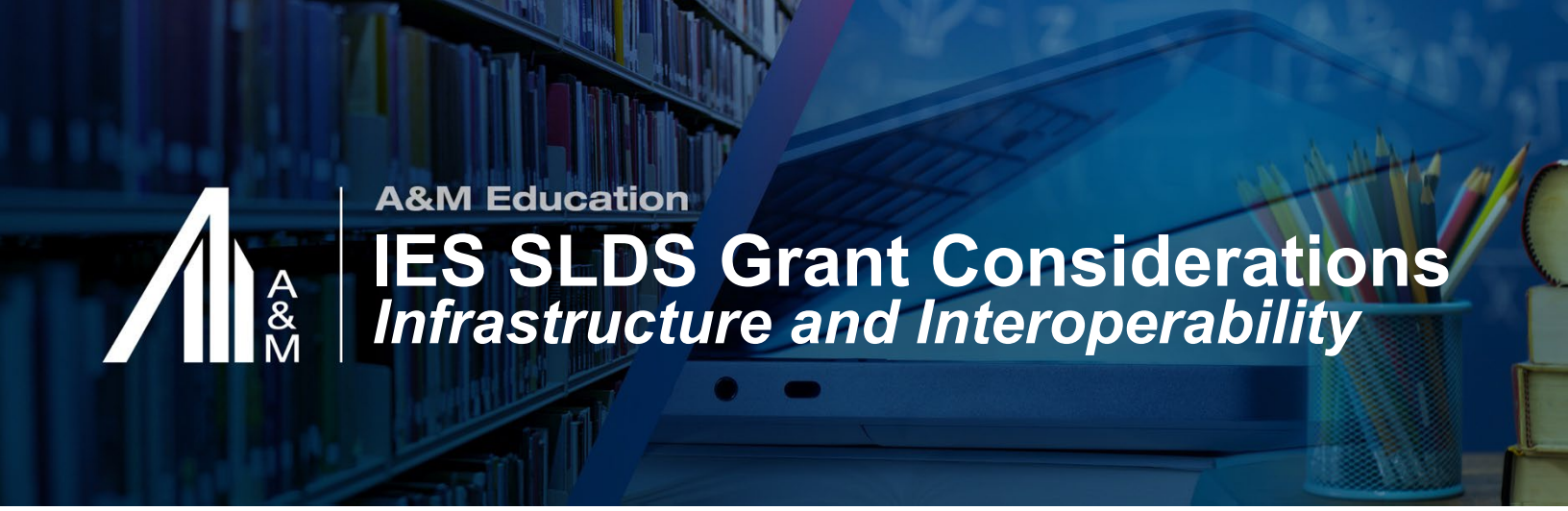


A&M Education

# IES SLDS Grant Considerations
## *Infrastructure and Interoperability*

*This paper is part of a series drafted in collaboration between Alvarez & Marsal Public Sector Services, LLC and the Council of Chief State School Officers in the interest of highlighting strategic considerations for state education agencies (SEAs) pursuing 2023 State Longitudinal Data Systems (SLDS) grants. They are organized based on the priority areas and key features of the 2023 application guidelines and address some major themes and challenges for states seeking to expand or develop their longitudinal data systems in alignment with the guidelines.*

## Problem Statement

Information in SLDS is largely focused on K-12 and postsecondary data. Content from the "bookends" – specifically early learning and workforce – is typically not well represented. In addition, data from adjacent systems such as health and human services or juvenile justice are rarely included. Interoperability and data linking remain major barriers to effectively acquiring, integrating and leveraging new data sets that can provide the insights needed to improve education outcomes for states.

## Grant Summary

Applicants seeking funding under this priority must describe how they would use the funds to develop or advance existing SLDS infrastructure in order to improve the linking and use of education data in the state. The focus of **Infrastructure and Interoperability** grants is on improving and expanding data linking and the interoperability of data across levels or sectors. In addition, the grant requires a state to use standards, such as Common Education Data Standards (CEDS), to demonstrate interoperability.

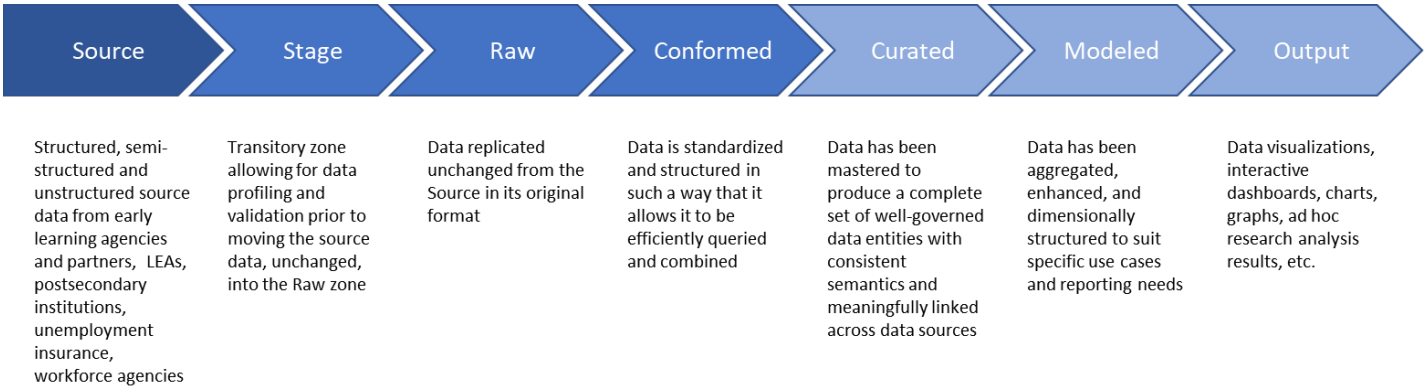
States cannot use this grant to support ongoing maintenance of current data systems, but the grant can be used to improve existing data systems to use data more effectively, create new systems where none existed or create new linkages between existing datasets.

At a minimum, states must propose to link at least one data source to currently existing K-12 data. Data sources may include early childhood education data, postsecondary data, workforce data or other non-educational data sources, including social services and juvenile justice data. States must also indicate how the proposed infrastructure investments will enable the state to provide data to stakeholders and other users to inform policy and practice.

## How Should States Think About This?

This grant focus area is intended to support infrastructure investments to enable an SLDS to acquire and link new data sources. The grant recognizes that infrastructure conceived and implemented more than a decade ago may no longer be adequate to meet new requirements, use cases or technologies needed to link and leverage new data with existing K-12 data. Rather than simply acquiring additional technologies to support linking new data sources, states can take this opportunity to fundamentally rethink their SLDS architecture and make infrastructure investments that align with this new architecture vision.

Modern data analytics architectures often leverage a “pipeline” or “zoned” pattern where data from sources is stored unchanged in “raw” form then “conformed,” “curated” and “modeled” to support use case driven outputs.




One key aspect of this architecture is that it *expects* data from sources to be in highly variable formats. Although some source data may be structured, other data may be semi-structured or unstructured. For example, data provided from a local education agency (LEA) may be structured in an Ed-Fi format. Postsecondary data may be structured in a PESC format. Unemployment Insurance (UI) data will likely be in a proprietary state agency format. Some advising data may take the form of unstructured notes.

In these architectures the source data is always stored in the data source original format, usually in a “data lake,” “delta lake,” or “lake house.” This allows the original source data to be reprocessed to create new outputs as use cases change, new data sources become available or source data structure changes over time. To accomplish this, data in varied formats needs to be transformed into a single normative form.

Many legacy data analytics architectures use an Extract Transform Load (ETL) approach to load data. Source data is extracted and transformed and then loaded into a database. A challenge with this approach is that any data that does not conform to the database structure is not captured. Modern architectures employ an Extract Load and Transform (ELT) approach where source data is stored in its original format and transformed as needed to support new use cases. The original source data is always available for future analysis.

Another key aspect is that data is semantically consistent. This means that the data “conforms” to a specific format and uses identical terms for data elements and values. Many legacy implementations using ETL will create an internal and proprietary structure to load and store data. Implementations using ELT can leverage a standard such as CEDS to help guide the design of the conformed data. This approach can improve interoperability and reduce the time and effort to transform data from source data standards. Organizations such as the T3 Network are working to provide tools and resources that map from individual standards into CEDS. Currently, the T3 Network’s tools are limited to postsecondary data sources, but this potentially represents a trend to develop mappings and tools to accelerate the process of conforming data and improving interoperability.

CEDS does not cover all potential domains for learner data. Data at the “bookends” of early learning and workforce and adjacent domains, such as health and human services, are underrepresented. However, CEDS covers a broad core of learner data, and if extensions to support additional data are well documented and managed, it can serve as a valuable tool to improve interoperability.

Another critical aspect of conforming is mastering and linking data. Early childhood education data, postsecondary data, workforce data, department of motor vehicles data or other non-educational data sources often will not share a common identifier for the learner. As attempts to link non-K-12 data expand, investing in robust mastering and linking tools becomes critical. In addition, this is an area where artificial intelligence (AI) or machine learning (ML) models could provide alternatives to more traditional master data management (MDM) tools.

Although the grant states that "at least one" new data source must be linked, states should target one or more new K12 data sources of sufficient complexity and disparity to justify new infrastructure investments. For example, if a state has a common student identifier, then adding postsecondary data from state institutions that utilize that same identifier may not be a compelling justification for improved data linking tools and processes.

Also, the target dataset must be "linkable." Said another way, there must be some feasible path to connect this data to existing K-12 data. Having a viable linking strategy and understanding any intermediate data that is required is critical.

Also, the data must be available to the entity hosting the SLDS. The grant mentions health and human services or juvenile justice data as a possible target data source. Accessing these data may be challenging in some states and require legislative approvals that would lengthen the timeline.

States should develop a long-term plan for managing the lifecycle of their data analytics pipeline sources and outputs. Without a lifecycle management plan, states run the risk of accumulating and maintaining data outputs that are no longer needed or used. States also risk developing many one-off data outputs directly from individual-level data sources instead of creating pre-built aggregations that allow states to quickly answer a wide variety of state policy questions within specific domains. The State Policy Questions paper discusses this topic in more detail.

In closing, implementing a modern data analytics architecture can speed time to development through leveraging off-the-shelf tools and services. States should develop a data analytics architecture vision for their SLDS and determine how a grant investment can be leveraged to make progress toward this vision.

## Exemplar Use Cases

### *Adding Workforce Data to an SLDS*

When sourcing workforce data for a SLDS:

1. **Individual level wage data is primarily sourced from UI records.** Access to UI data will require data sharing agreements with the state agency that manages UI. This may require state executive and/or legislative action. Linking UI data to individual learner records requires a mastering and linking effort. Many states utilize DMV records to provide a path to link K-12 and postsecondary data to workforce data as DMV data contains both personal attributes (gender, ethnicity and physical address) and social security number (SSN), which is included in UI data.
2. **Additional sources of wage data are available but may not support specific use cases. Examples include:**

| Source | LinkedIn | Lightcast (Formerly EMSI Burning Glass) | Equifax | SWIS (State Wage Interchange System) |
|---|---|---|---|---|
| **Type of Data** | Labor Market | Labor Market | Employment History<br>Wage<br>Other Personal | Wage (UI) |
| **Individual-Level** | No (aggregate only) | No (aggregate only) | Yes | Yes (Anonymized) |
| **Accessible** | Yes | Yes | Yes | Limited to WIOA program participants |
| **Affordable** | $$$ | $$$ | $$ | N/A |

3. **Multistate data is challenging.** States that have large metro areas that span neighboring states (e.g., New York, New Jersey, Washington DC, Virginia, Maryland and many northeastern states) will need to source UI data from multiple states to get a full picture. Developing outcome measures for secondary and postsecondary learners is also challenging when they leave the state for employment.

## Questions to Consider

What questions can you ask to help determine where to start and the level of infrastructure investment that may be required to modernize your SLDS and successfully link new source datasets?

1. **What is the architectural model of your current SLDS?** How is data transformed and loaded into your SLDS? Are the original data formats preserved and are they accessible to your data analytics tools? Is an ETL or ELT transformation approach used? Is your data hosted by a cloud provider? Is hosting your SLDS on a single cloud provider acceptable or is cloud portability a requirement? Can you support unstructured or semi-structured data?

2. **What additional data sources are available to link?** What additional data would allow you to answer critical policy questions? What constraints are present for gaining access to new data to link? Will executive or legislative action be necessary to gain access to that data?

3. **How much overlap is there between your existing SLDS data and the new data you are targeting?** After close examination, many potential data sources only offer a small percentage of overlap with existing K-12 and postsecondary data. Linking data with only marginal overlap may not provide enough additional benefit to offset the implementation expense and effort.

4. **If your organization lacks expertise in modern data analytics architectures, are there resources to help?** Are there other agencies in the state with this experience? Are there potential partnerships with universities that may have this expertise?

5. **What about historical data?** Converting historical data into a modern architecture can be complex and costly. Does the original source data exist? Are the data and transformations made on it well understood? How much historical data is necessary to answer current policy questions?

## KEY CONTACTS`


**Erin Kenny**
Managing Director
+1 415 690 3708
ekenny@alvarezandmarsal.com

**Paul Tearnen**
Managing Director
+1 714 785 3706
ptearnen@alvarezandmarsal.com

**Eli Pristoop**
Director
+1 206 753 0857
epristoop@alvarezandmarsal.com


Follow A&M on:



### ABOUT ALVAREZ & MARSAL

Companies, investors and government entities around the world turn to Alvarez & Marsal (A&M) for leadership, action and results. Privately held since its founding in 1983, A&M is a leading global professional services firm that provides advisory, business performance improvement and turnaround management services. When conventional approaches are not enough to create transformation and drive change, clients seek our deep expertise and ability to deliver practical solutions to their unique problems.

With over 7,500 people providing services across six continents, we deliver tangible results for corporates, boards, private equity firms, law firms and government agencies facing complex challenges. Our senior leaders, and their teams, leverage A&M's restructuring heritage to help companies act decisively, catapult growth and accelerate results. We are experienced operators, world-class consultants, former regulators and industry authorities with a shared commitment to telling clients what's really needed for turning change into a strategic business asset, managing risk and unlocking value at every stage of growth.

When action matters, find us at: **alvarezandmarsal.com**

ALVAREZ & MARSAL
LEADERSHIP. ACTION. RESULTS.