

# FROM DATA TO CLASSROOMS

## THE TEACHER WORKFORCE ROADMAP

Prepared By  
**SETA-ED**

Date  
**FEBRUARY 2026**

Co-Published By  
**BEYOND100K**

**NEW MEXICO PUBLIC EDUCATION  
DEPARTMENT (NMPED)**

**ARKANSAS DEPARTMENT OF EDUCATION,  
DIVISION OF ELEMENTARY AND  
SECONDARY EDUCATION**

# INTRODUCTION

Through the “**Solution Lab to Accelerate Teacher Workforce Data Systems,**” Beyond100K is helping states to build better tools and processes in order to proactively get ahead of and rapidly respond to teacher shortage crises. To do this, states require the ability to accurately forecast trends in the workforce, and dynamically create the conditions to incentivize, recruit, and prepare the workforce to meet the future needs of the economy and our students. Through convening with states all across the nation to discuss the major barriers and bright spots to solving teacher shortages, one key driver became apparent: robust data systems. Workforce data systems collect data across the full educator pipeline and make that data discernible and accessible, providing a clear picture of the current supply and demand of the teacher workforce. A centralized data system enables better forecasting and workforce planning to address, get ahead of, and ultimately avoid localized and widespread teacher shortages.

Beyond100K joined forces with research partner, SETA-ED (Scientific and Engineering Technical Assistance for Education), to conduct a 6-month field scan to assess the current state of teacher workforce data systems and identify opportunities to develop guidance for states in enhancing the capacity, utility, and effectiveness of their systems. SETA-ED interviewed 11 states and 8 stakeholder organizations that work deeply on the challenges relating to measuring and solving teacher shortages at the state-level. What was discovered during the field scan work was: not one state has an entirely perfect workforce data system from end-to-end, but many, if not all, states have implemented one or more components of a robust teacher workforce data system.

To guide states in developing or enhancing their teacher workforce data systems, we have devised a “**Roadmap for Developing a Robust Teacher Workforce Data System**” at the state level, consisting of 10 critical steps for building robust and highly effective data systems. While these steps could apply to many different data systems and use cases, the intention of this Roadmap is specifically focused on helping states measure and track K-12 teacher supply and demand.

## WHY DOES THIS MATTER?

Teaching is a profoundly rewarding but also hard profession. Through Beyond100K's extensive mapping process, the [Beyond100K network uncovered over 100 challenges](#) that prevent STEM teachers from entering or persisting through the profession, including insufficient compensation, unsupportive workplace conditions, or even external factors such as housing costs and lengthy commutes can leave teachers feeling unequipped for success and disenchanted with the profession. Some quality candidates never become teachers because the pathways to becoming teachers are not always supportive of diverse candidates or candidates coming from non-traditional preparation programs. This challenge demonstrates how even well-intentioned systems can create barriers that unevenly impact marginalized populations if not carefully implemented and monitored. When widespread, these issues lead to teacher shortages.

**How does this impact students?** High rates of teacher turnover and substitute teacher use, which are common conditions attributed to teacher shortages, negatively impact student learning and success. Students thrive the most in consistent environments with high-quality teachers who are specifically trained for their position. If we want our students to succeed academically, we must prioritize building and retaining a highly-skilled and evenly distributed teacher workforce.

There is no shortage of challenges to be addressed, but state-level resources are often limited and in high demand, so how do states identify the most important issues to tackle first? Identifying the highest priority issues and putting forward solutions to address them requires significant coordination and evidence building. Our goal is to help states pinpoint the most critical barriers impacting teacher recruitment and retention, and build rich evidence and coordination across stakeholders to solve those challenges. A visual way to think about this is: what if every person in the state trying to become a teacher got in line and we talked to them? What would be the most significant and common set of challenges they're facing (e.g., is it financial, is it licensure requirements, is it logistical barriers), and what could the state do to alleviate those recruitment issues? What if we did the same with currently staffed teachers or teachers who have recently left the profession in order to better understand retention? Many of the challenges we would identify are not unique or uncommon, and would greatly benefit from being addressed through state-level policies and initiatives (e.g., funding support for teacher salaries, efforts to improve workplace conditions, or initiatives to increase affordable housing). While working with those closest to the challenge yields the best solutions, we need a way to do this at scale.

## REALITY CHECK

Most states already collect the data that is needed to assess their most pressing challenges. The data are scattered across EPP records, HR systems, licensure databases, and district reports. The challenge isn't generating new data—it's bringing together what exists and making it usable. Before incurring any additional burdens on school districts, educator preparation programs (EPPs) or other institutions affiliated with the teacher pipeline, we want to fully utilize what already exists. Bringing data together and making it usable and useful, not only unlocks states' abilities to identify key challenges and barriers across their teacher pipeline, but also enables them to identify and amplify areas and initiatives that are successful in bringing about sustainable change.

## PURPOSE AND HOW TO APPROACH THE ROADMAP

This Roadmap is designed to help states build a robust teacher workforce data system, but what does this look like? An effective teacher workforce data system should (1) enable states to rapidly identify current trends in teacher supply and demand, (2) proactively forecast future needs to ensure states are preparing today the teachers of tomorrow, and (3) catalyze states' ability to drive informed change through state-level policy or other forms of capacity building. States need to both understand the nature of their current workforce capacity and forecast how that capacity may or may not meet the needs of the future. Without both of those perspectives, well-meaning strategies to better the workforce may wind up ineffective in either the near term or the long term.

In an effort to traceback a path to achieving this common goal, we have identified a set of key components that are critically important when developing or improving a state's teacher workforce data system. These components span both operational and technical aspects to standing up and managing a workforce data system. For each component, we have highlighted examples, referred to as bright spots, and provided resources to help guide states in implementation.

The states successfully tackling the teacher workforce challenge are using data to answer critical questions faster, such as:

- Why are new teachers leaving the highest-need schools?
- Which incentives actually improve retention?
- Where and how should states invest in alternative certification programs?

**Every component in this Roadmap exists to help answer questions like these, and take action on the answers.**

## CHOOSE YOUR OWN ADVENTURE

If starting from scratch, these components may be done in order, however, most states already have one or more of these components in place, and therefore the Roadmap encourages states to choose their own adventure! State leaders should select which components matter most to them right now and evaluate whether each component is readily attainable or requires significant effort for their state to implement. This Roadmap is designed to be a general starting point—please keep in mind the components will need to be adapted, customized and implemented according to each state's specific needs and operating practices.

***Companion Guides** are provided in the **Appendix** for state leaders to assess the current status of their state data systems across the 10 key components, and navigate where to start and which items to prioritize. The Companion Guide includes a self-assessment checklist, progress rubric, reflection prompts, and planning tables for next steps. We recommend printing or creating a digital copy of the Companion Guide to fill in as you read through the Roadmap.*

# OTHER FORMS OF SUPPORT

## ■ **Bootcamp Series**

The Roadmap was supported by a **Bootcamp Series** that took place throughout the Fall of 2025. The Bootcamps were opportunities for states to receive tailored instruction and support from Beyond100K, SETA-ED and policy and technical advisors to discuss ways to customize and apply these components of this Roadmap into practice. Through the Bootcamps, states were able to hear directly from bright spot states and organizations on what specifically has worked for them and how.

## ■ **Hands-On Support**

Beyond this Roadmap and Bootcamp Series, Beyond100K and SETA-ED want to support you in your data systems—from design through implementation. Our teams provide various supports to states, including policy guidance, data privacy counsel, strategic communications, and hands-on technical and engineering support. For additional information and support, please reach out to [info@seta-ed.com](mailto:info@seta-ed.com) or [meghan@beyond100k.org](mailto:meghan@beyond100k.org).

# THE ROADMAP

## THE PROMISE: FROM DATA TO IMPACT

Imagine knowing exactly where your state's teacher shortages will emerge next year, and having the tools to prevent them.

Illinois did this using comprehensive vacancy data to target support to their most understaffed districts. The result? Over 51% grant-receiving districts saw a decrease in unfilled positions in year one alone. They didn't just react to shortages; they predicted, prevented, and measured their success.

Texas transformed compliance into improvement by turning mandatory data collection into actionable intelligence. Now, over 120 educator preparation programs across the state have access to 300+ data elements at no cost to inform and strengthen their programs.

These examples demonstrate the power and promise of a robust teacher workforce data system.

## WHAT YOU'LL GAIN

This Roadmap helps you build three critical capabilities:

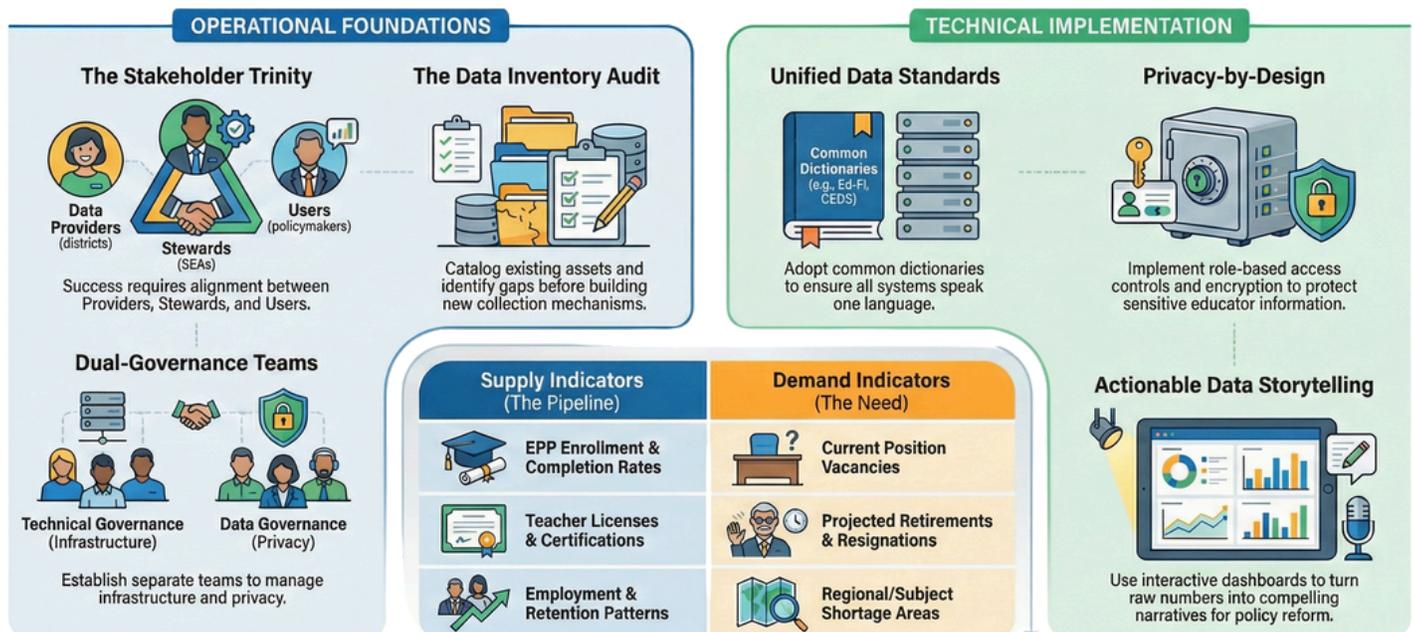
1. **See clearly:** rapidly identify current trends in teacher supply and demand across your state and answer critical questions on the state of your teacher workforce
2. **Plan ahead:** proactively forecast future needs before shortages emerge
3. **Drive change:** make informed policy decisions backed by evidence, not guesswork

# YOUR QUICK-START PATH

- If you have 5 minutes:** Review the visual roadmap overview (below) to quickly see all the components and how they connect.
- If you have 30 minutes:** Read Components 1 (Stakeholder Buy-in) and 2 (Data Discovery) to assess your current state.
- If you're ready to build:** Components 3-7 provide technical implementation guidance with real examples from states.
- If you want to prove value:** Jump to Components 8-9 to see how dashboards and data storytelling drive policy impact.

## A Roadmap to Robust Teacher Workforce Data Systems

A high-level blueprint for state leaders to transition from siloed data to integrated systems, aligning operational support with technical infrastructure to measure and address teacher shortages



## READY TO START?

Pick the component that addresses your biggest pain point today. **You don't have to be perfect, you just have to start, and we're here to help!**

# TABLE OF CONTENTS



<b>10</b>	<b>Component 1: Stakeholder Buy-in</b>	
	1.1 Identify Stakeholder Groups	10
	1.2 Define Requirements for Each Stakeholder Group	12
	1.3 Develop a Stakeholder Plan of Action	13
<b>14</b>	<b>Component 2: Data Discovery</b>	
	2.1 Inventory Existing Data	14
	2.2 Identify Gaps and Prioritize Next Steps	15
<b>18</b>	<b>Component 3: Building a Team</b>	
	3.1 Establish a Technical Team	18
	3.2 Establish a Data Governance Team	20
<b>23</b>	<b>Component 4: Standardizing Data</b>	
	4.1 Build a Data Dictionary	24
	4.2 Define Key Terms	27
<b>30</b>	<b>Component 5: Data Infrastructure Implementation</b>	
	5.1 Design and Build the Architecture	30
	5.2 Specify Data Security and Privacy Requirements	33
<b>36</b>	<b>Component 6: Data Collection Mechanisms and Intervals</b>	
	6.1 Develop Clear and Easy Data Collection Protocols	36
	6.2 Determine Data Collection Intervals and Format	37
<b>39</b>	<b>Component 7: Clean, Quality Check, and Merge Data</b>	
	7.1 Data Validation and Quality Assurance Processes	40
	7.2 Link and Join Data from Multiple Systems	42
<b>45</b>	<b>Component 8: Data Visualization and Reporting</b>	
	8.1 Dynamic Dashboards and Data Visuals	45
	8.2 Automated Static Reporting	49
<b>51</b>	<b>Component 9: Supporting Data-Driven Initiatives</b>	
	9.1 What is Data Storytelling?	51
	9.2 Identify and Support Data-Driven Policy Initiatives	52
<b>54</b>	<b>Component 10: Sustainability and Expansion</b>	
	10.1 Engage Stakeholders on Ways to Leverage Data and Gather Feedback	54
	10.2 Making Improvements and Continuing to Modernize the System	55
<b>58</b>	<b>Closing</b>	
<b>59</b>	<b>Acknowledgements</b>	
<b>60</b>	<b>Glossary</b>	
<b>64</b>	<b>Appendix</b>	

# COMPONENT 1: STAKEHOLDER BUY-IN

## OPERATIONAL

—The operational support system that enables the data system to exist and flourish

Building a data system that enables a state to accurately measure and respond to teacher shortages demands a strong foundation of stakeholder engagement that enables the system to thrive. This foundational component focuses on identifying and securing commitment from each of the key players in the teacher workforce ecosystem. Those key players include educator preparation programs (EPPs) and school districts who provide critical sources of data; state education agencies (SEAs) that steward the data; and policymakers and researchers who rely on the data to drive policy initiatives and decision-making. Success hinges on bringing together this complex web of Data Providers, Data Stewards, and Data Users across the state, understanding their unique needs and constraints, and establishing clear protocols for how these diverse stakeholders will collaborate to collect, share, analyze, and leverage teacher workforce data. This component emphasizes the critical importance of early and sustained engagement with all stakeholder groups, addressing their concerns about usability, data privacy and security, and creating mechanisms for ongoing input and feedback. By establishing strong stakeholder buy-in and governance frameworks from the outset, states can ensure their teacher workforce data system will not only function effectively but will also be embraced and utilized by the very communities it is designed to serve.

## 1.1 IDENTIFY STAKEHOLDER GROUPS

**Objective:** Identify the key stakeholders and achieve stakeholder buy-in and a willingness to work together to improve teacher workforce data systems across the state.

### Who are the key stakeholder groups?

When it comes to collecting, managing, and utilizing data indicating teacher supply and demand, the most important stakeholders to identify are divided into three main groups: Data Providers, Data Stewards, and Data Users.

## ■ Data Providers

are those who contribute data to the system. Data Providers include agencies that collect some aspect(s) of teacher supply and demand data through their standard operating practices, and provide that data to the state. Data Providers may include: EPPs (e.g., higher ed institutions, residency or apprenticeship programs), teacher license boards who define requirements for teacher licenses, 3rd-party vendors

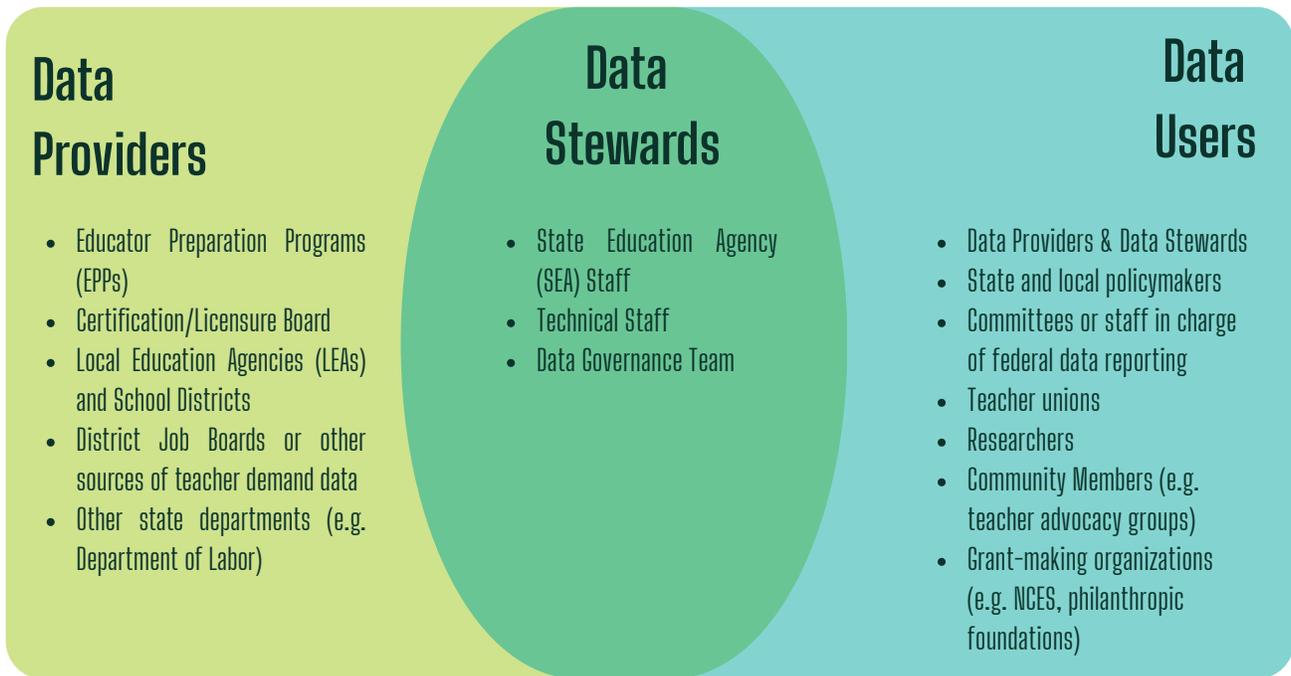
## ■ Data Stewards

collect, join, and manage centralized data stores, or manage systems on behalf of an education agency (e.g., Managed Service Providers). Data Stewards might serve as operational staff, data analysts, or technical support staff. Data Stewards define the rules and operating procedures for interacting with the centralized data system. The Data Providers and Data Stewards work together to determine how the data must be governed, especially how personal data shall be protected and secured, and in what form data may be shared with other stakeholders. Data Stewards include SEA and technical staff who manage the data system (e.g., IT staff, data engineers, system administrators).

## ■ Data Users

usually include both of the prior groups (Data Providers and Data Stewards) plus additional folks who leverage data to inform data-driven research and/or policy initiatives (e.g., external partners, researchers, or policymakers). These users may view dashboards, pull datasets for analysis, and develop or review downstream reports. Their efforts include using data to develop policies or programs that aim to reduce teacher shortages and bolster the teacher workforce across the state. Furthermore, Data Users identify what information is needed to solve teacher shortage challenges in the state.

# Key Stakeholders



## 1.12 DEFINE REQUIREMENTS FOR EACH STAKEHOLDER GROUP

**Objective:** Identify what each stakeholder needs in order to effectively contribute to and/or leverage the data system (e.g., What access points are needed to submit data? What options are available for querying data? What specific data elements are needed to support policy initiatives?)

### Where to Start?

Once the state has identified who makes up each of the three stakeholder groups, it's helpful to think through the mechanisms and timelines in which these stakeholders will need to work together to collect, share, analyze, and effectively leverage the teacher workforce data system. Some questions to ask may include:

- What kinds of data might each of these stakeholders need to leverage, and how often?
- How will they be able to access data, and what is required of them in order to find and access data?
- What data security, privacy, or legal measures may need to be put into place for different types of data (e.g., raw data, PII, de-identified vs. aggregate data)?
- How often and through what opportunities will each of these stakeholders have to weigh in on the development and utility of the state's teacher workforce data system?

## 1.3 DEVELOP A STAKEHOLDER PLAN OF ACTION

**Objective:** Develop a plan of action to engage stakeholders at the right level of support to ensure the utility and success of the data system.



### Activity: Develop A Stakeholder Engagement Plan of Action

Identify members of each of the key stakeholder groups in your state, and develop a plan of action to engage the stakeholder groups at a level that will ensure the state's teacher workforce data system will be successful and maximally useful to all stakeholders. A plan of action for engaging stakeholders might look like: [Stakeholder Plan of Action](#)

### BRIGHT SPOT EXAMPLE



**[The Illinois Workforce and Education Research Collaborative \(IWERC\)](#)** excels at uniting key stakeholders around data and data-driven education research initiatives statewide. IWERC conducts research across early childhood, K-12, and postsecondary education, analyzing its intersections with job-readiness. This research effectively connects diverse statewide contexts, particularly rural and urban areas, and stakeholders at all levels, from

practitioners to policymakers. IWERC's core practice is to establish deliberate partnerships from the outset, including with other researchers, research participants, and potential end-users of research. Modeled after other research-practice partnerships, such as the Consortium on School Research at the University of Chicago, IWERC's agenda is shaped by Illinois stakeholders. IWERC collaborates with state agencies, school districts, early childhood programs, colleges, workforce development initiatives, and other educational partners to research and resolve critical issues.

# COMPONENT 2: DATA DISCOVERY

OPERATIONAL

## —Identifying what we know and what we don't know

Before building or enhancing a teacher workforce data system, states must first understand what data they already possess and identify critical gaps in their information landscape. This discovery component serves as a comprehensive audit of existing data assets, creating an inventory of all teacher supply and demand data currently collected by the SEA. The process involves cataloging not only what data elements are gathered, but also documenting where the data are stored, in what format it exists, and how frequently it is updated. A crucial aspect of this component is mapping existing data elements to specific indicators of teacher supply and demand, such as EPP enrollment rates, licensures, placements, and position vacancies. Through this systematic discovery process, states can pinpoint knowledge gaps, including areas where critical teacher workforce indicators are not being tracked or where data exists but remains inaccessible for policy and planning use. The discovery process can also highlight where measurement inconsistencies and definitional challenges create barriers to comprehensive analysis. This foundational understanding enables states to make informed decisions about what new data efforts (e.g., collection, standardization, accessibility, visualization, or analysis) the state should focus on and which stakeholders must be involved to fill these gaps.

## 2.1 INVENTORY EXISTING DATA

**Objective:** Catalog the existing state-level data assets pertaining to teacher workforce, and make the format for cataloging this data extensible so it can continue to be added to it as the state expands data collection indicative of the teacher workforce.

**Use the following steps** to discover and inventory the state’s existing teacher supply and demand data:

1. Pull from existing resources and institutional knowledge to develop a comprehensive list of the data elements that are currently being collected.
  - a. *Note:* If you’re struggling with where to start, an easier starting point might be cataloging the baseline data elements every state gathers to meet federal data reporting requirements (e.g., Title II Reports). Then, consider beyond those data elements for what additional indicators of teacher supply and demand does the state also collect for other purposes?
2. Document where the data are stored (once collected by the state), the source system (where those data originate), the structure of the data, and on what scheduled intervals those data are collected.
  - a. Data storage: database or file system managed by either the state or a managed service provider (MSP), cloud-based storage systems, or physical data storage systems (e.g., servers, storage drives)
  - b. Source systems: district student information systems, HR management systems, certification/licensure databases, EPP records, etc.
  - c. Data type/structure: relational database queries, flat files (e.g., CSV, Excel), specific application data formats, manual records, or paper-based systems
3. If different from the source system, identify the agency, department, or organization that maintains authoritative control and ownership over the data (i.e., the Data Steward or Data Provider)

## 2.2 IDENTIFY GAPS AND PRIORITIZE NEXT STEPS

**Objective:** Align existing data assets with indicators of teacher workforce supply and demand, identify gaps, and prioritize areas to address with the highest estimated returns on investment.

**Use the following steps** to pinpoint data accessibility and information gaps that limit the state’s ability to understand and address teacher shortage problems:

1. **Map data elements to indicators of teacher supply and demand** by crosswalking the data elements identified through data discovery with key indicators for assessing teacher workforce dynamics.

# Sample Indicators of Teacher Supply & Demand

## Supply

- EPP Enrollment and completion rates
- Teacher licenses and certifications
- Employment data (e.g. hiring, assignments, retention, attrition, re-entry patterns)
- Current teacher workforce demographics (age, experience, qualifications) by geographic region and subject area



## Demand

- Current position vacancies and unfilled roles
- Projected teacher retirements and resignations
- Class sizes, student-teacher ratios, and educator caseloads
- State-designated teacher shortage areas by subject matter and geographic region
- Emergency permits and temporary authorizations

**2. Identify knowledge and data gaps:** determine what teacher supply and demand indicators are missing from current data collection efforts and would enhance the state's awareness with regard to teacher shortage challenges. The following are some examples of common challenges we heard from states:

- a. Supply-side gaps: missing data on teacher recruitment pipeline, such as quality indicators across educator preparation programs, teacher placement after certification, projections for pending licensed teachers in the pipeline, teacher job satisfaction and intent to stay, competitiveness analysis when compared to other states' teacher working conditions, incentives, and benefits
- b. Demand-side gaps: absence of predictive models for future teacher needs, limited vacancy data or data on hard-to-staff positions beyond state shortage designations, insufficient information on workload and working conditions that drive demand

**3. Collaborate with stakeholders:** revisit 1.3: Developing a Stakeholder Plan of Action and identify which Data Providers, Stewards, and Users can contribute to filling identified gaps through enhanced data collection, sharing agreements, or other collaborative data initiatives.

4. **Establish priority and feasibility:** evaluate the practical viability of collecting missing data elements (considering stakeholder capacity and system limitations) and associated return on investment. For newly desired data collections, prioritize collecting data that will yield the highest impact, and minimize any additional burden on Data Providers.



- Template for inventorying existing data, and assessing gaps in teacher supply and demand: [Data Discovery & Gap Analysis: Blank Template](#)
- Template User Guide: [Data Discovery & Gap Analysis User Guide](#)

# COMPONENT 3: BUILDING A TEAM

OPERATIONAL

—Assembling the human capital needed to build and run the system

A successful teacher workforce data system requires more than just technology and data—it demands a skilled team responsible for designing, implementing, governing, and maintaining the data infrastructure. This component focuses on assembling two critical teams that will serve as the backbone of the data system: a technical team responsible for the technological infrastructure and a data governance team responsible for managing stakeholder relationships and ensuring responsible data management. The technical team brings essential expertise in system architecture, database management, cybersecurity, and software development. The technical team helps states navigate crucial decisions about whether to build custom solutions, purchase existing platforms, or leverage shared resources. Meanwhile, the data governance team serves as the operational heart of the system: establishing and maintaining relationships across diverse stakeholder groups, developing data sharing agreements, and ensuring that the system meets the evolving needs of Data Users while maintaining strict privacy and security standards. These teams must work in close coordination, with the technical team providing the infrastructure capabilities and the governance team ensuring that the system serves its intended purpose of improving teacher workforce planning and policy development. Given the specialized nature of this work and the limited technical capacity within many SEAs, states must carefully consider whether to build internal capacity, partner with managed service providers (MSPs), or pursue hybrid approaches that combine internal leadership with external technical expertise.

## 3.1 ESTABLISH A TECHNICAL TEAM

**Objective:** Build or outsource a team of technical staff to help design, implement, and maintain the technical data infrastructure needed for a robust teacher workforce data system.

## Overview

The technical team serves as the technological backbone of the data system, bringing specialized expertise in system architecture, database management, cybersecurity, and software development. This team may consist of internal SEA staff if technical capacity is available, or may be outsourced to a managed service provider (MSP) or technical consultants, depending on the state's resources and strategic preferences.

## Key Responsibilities:

- System architecture design: develop comprehensive technical specifications for the data infrastructure in order to provide key functionality, including data collection, storage, processing, and dissemination—see Component 5 for more details on key architecture decisions.
- Build vs. buy vs. borrow: inform strategic decisions about whether to develop custom solutions, purchase from existing platforms, or leverage shared resources from other systems.
- Security implementation: establish and maintain cybersecurity protocols, including data encryption, access controls, system monitoring, and incident response procedures
- Integration management: enable seamless flow of data between various source systems, databases, and user-facing applications.
- Performance optimization: monitor system performance, troubleshoot technical issues, and implement improvements to ensure reliable and efficient operations.
- Technical documentation and user support: create and maintain user guides and system documentation, such as data dictionaries, API specifications, and operational procedures, and oversee user access management and support.

## Technical Staffing Options:

- **Internal Technical Staff:** utilize existing state IT resources
  - Benefits: direct control, institutional knowledge, deep understanding of state-specific needs
  - Requirements: sufficient technical expertise and available ongoing capacity
- **Managed Service Providers (MSPs):** partner with specialized vendors for complete system management
  - Benefits: proven expertise, scalable resources, comprehensive support
  - Considerations: vendor selection, contract management, familiarity with applicable data compliance frameworks

- **Technical Consultants (Hybrid Approach):** engage specialists for specific implementation phases or components
  - Benefits: targeted expertise, balance control and capability, leverage specialized vendor capabilities, flexible engagement, knowledge transfer opportunities to internal teams for long-term sustainability
  - Applications: system architecture, security implementation, integration development



**SETA-ED** (Scientific and Engineering Technical Assistance for Education) provides white-label technical and engineering consulting services, including end-to-end data infrastructure implementation and support. SETA-ED offers a robust team of data engineers, cloud architects, software engineers, IT and cybersecurity specialists. To learn more about the services and solutions offered, reach out to [info@seta-ed.com](mailto:info@seta-ed.com).

Ed-Fi Alliance's "[Best Practices for Coordinating with Technology Providers](#)" outlines key strategies for state education agencies to effectively collaborate with technology vendors for outsourcing technical and engineering expertise. (Note: this resource is specific to implementing the Ed-Fi Data Standard and Technology Suite—discussed in detail in Component 4: Standardizing Data.)

## 3.2 ESTABLISH A DATA GOVERNANCE TEAM

**Objective:** Create a team to manage and govern the state's teacher workforce data system, ensuring the system meets stakeholder needs while maintaining standards for data privacy, security, and quality.

### Overview

The data governance team serves as the operational heart of the teacher workforce data system, organized by SEA or a specific department focused on teacher preparation and retention. This team may take various forms depending on the state's structure and preferences, such as a coalition of stakeholders, data privacy advisors, SEA staff, technical support staff, or MSPs.

### Key Responsibilities:

- Stakeholder liaison and coordination: serve as the primary communication bridge between Data Providers, Data Stewards, and Data Users, and identify opportunities to enhance data collection, sharing, and utilization across the teacher workforce ecosystem.
- Legal and policy framework development: establish memoranda of understanding (MOUs) and data sharing agreements (DSAs) across stakeholder groups to establish standards for secure and compliant data sharing, and develop comprehensive policies for data governance, including privacy protections, security controls, and appropriate use guidelines—ensuring compliance with federal and state privacy laws.
- Data needs assessment and gap analysis: continuously assess what information Data Users need to support their teacher workforce initiatives and policy decisions, and evaluate whether Data Providers already possess needed information that simply requires improved sharing protocols vs cases where Data Providers need additional capacity, tools, or support to collect and provide required information, and coordinate with the technical team to design and implement solutions.
- Data quality and stewardship: ensure each critical data domain (e.g., certification, employment, educator preparation) has a dedicated Data Steward or Data Governance team to facilitate data submissions and ensure data validation and quality control measures—for additional guidance, see section **7.1: Data Validation and Quality Assurance Processes**.

**What are the common challenges?** Below are some of the pitfalls SEAs have identified in forming the technical and data governance support needed to support a robust teacher workforce data system:

- Resource constraints: SEAs often have very limited human resources, capacity, and technical expertise needed to drive comprehensive data initiatives.
- Political support: strong political leadership can catalyze data-driven initiatives, making it easier to secure adequate funding and resources for data-centric initiatives.
- Stakeholder coordination: managing diverse stakeholder groups with varying priorities, technical capabilities, and institutional cultures requires skilled facilitation and relationship management.
- Sustainability: ensuring long-term system viability requires ongoing commitment and resources—spanning multiple administrative cycles and budget periods.



## KEY RESOURCES

The following are two example templates for generating a Data Sharing or Privacy Agreement:

- [Model Data Sharing Agreement](#)
- [National Research Data Privacy Agreement \(NRDPA\)](#)



## BRIGHT SPOT EXAMPLE

The [Maryland Longitudinal Data System \(MLDS\) Center](#) exemplifies how strategic coalition-building and sustained political commitment can create a powerful, centralized data infrastructure for education workforce planning. Established by state law in 2010 and operational since 2013, the Center operates as an independent unit of state government overseen by a 15-member governing board, with partnerships spanning six state agencies, including the

Department of Education, Higher Education Commission, and Department of Labor. What makes Maryland's approach particularly effective is their unique shared staffing model, where MLDS funds liaison positions with partner agencies, enabling staff to contribute half-time to the Center and half-time to the partner agency. This shared staffing model has helped to bolster institutional buy-in, continued communication, and close partnership across agencies. This collaborative structure, combined with university partnerships for research services and a deliberate strategy to bring system development and maintenance in-house, has created a sustainable, cost-effective solution that can dynamically respond to changing workforce demands. The Center's success lies not just in its technical capabilities, but in its ability to align diverse stakeholder interests around a shared vision of data-driven education and workforce planning.

# COMPONENT 4: STANDARDIZING DATA

## TECHNICAL

### —Creating a common language around teacher workforce information

One of the most critical yet challenging aspects of building a comprehensive teacher workforce data system is establishing standardized definitions, formats, and coding schemes. Without this, data integration and analysis across diverse data sources is nearly impossible.

Below are examples of the common challenges identified with states relating to data standardization through our Field Scan, which we used to inform the development of the Roadmap. Many states experience:

- Inconsistent nomenclature and data definitions (e.g., varying definitions for “vacancy,” “teacher shortage,” “qualified teacher”)
- Differing or inconsistent methods for measuring key indicators of teacher shortages (e.g., vacancy rates, retention rates, attrition rates)
- Timing discrepancies (e.g., different reporting periods, fiscal vs. academic year misalignments)
- Missing standardized identifiers or common codes
- Incomplete historical records or limited ability to track data longitudinally
- Variations in validation procedures, error checking standards, and data cleaning protocols across data sources

To begin to address these challenges, this component builds on and standardizes the outputs from task 2.1: Inventory Existing Data. The component focuses on utilizing a data standard to develop a data dictionary and establish shared language. Data standards address the fundamental problem that different stakeholder organizations often use slightly different definitions for the same concepts, collect data in incompatible formats, and employ different measurement methodologies—creating barriers to comprehensive workforce analysis and cross-system data sharing. For instance, without standardized data elements, states cannot link information from EPPs, school district HR systems, certification databases, or other sources—making integrative analytics across these various data sources almost impossible.

This standardization work requires developing a comprehensive data dictionary that provides consistent definitions for key concepts, establishes common coding schemes for critical data elements, and clarifies differing business rules or measurement approaches for complex indicators. The process involves not only technical decisions about data formats and structures, but also the more challenging work of building consensus among stakeholders about how to define and measure important concepts such as teacher shortages, position vacancies, teacher qualifications, or teacher effectiveness. By establishing these standards early in the system development process, states can ensure that their teacher workforce data will be linkable across time periods, geographic regions, and organizational boundaries, ultimately enabling robust integrative analysis.

## 4.1 BUILD A DATA DICTIONARY

**Objective:** Leverage a common data standard in building a comprehensive data dictionary to define key variables, data formats, coding schemes, and measurement methodologies.

### Overview

A comprehensive data dictionary serves as the technical foundation for data standardization, providing detailed specifications for how each data element should be defined, formatted, coded, and collected. This dictionary should align with established education data standards while addressing the state and district-specific modifications to defining and collecting key data elements.

## Implementation Steps:

The following is a set of non-prescriptive steps to building out a data dictionary:

1. **Adopt a Data Standard:** The two most robust and commonly used open education data standards are the Common Education Data Standards (CEDS) and the Ed-Fi Data Standard. These two standards are designed to serve as stand-alone options, but they can also be complementary to one another and implemented in tandem.
  - a. **CEDS:** standardizes definitions, formats, and relationships for education data elements
  - b. **Ed-Fi** Data Standard: a specification for data exchange and interoperability, developed and maintained by the Ed-Fi Alliance (please see the Glossary for more information on the Ed-Fi Alliance)
2. **Specify the dimensions and level of disaggregation for each variable:**
  - a. Geographic granularity: state, district, or school-level data
  - b. Subgroups and demographics: such as race/ethnicity, school type (public, private, charter), grade levels, subject areas, region type (urban, suburban, or rural), or economic status (low, medium, or high income area)
3. **Record collection schedules and variability:**
  - a. Collection intervals: real-time updates, daily feeds, weekly, monthly or quarterly snapshots, annual collections, ad-hoc requests, event-triggered updates (e.g., upon hiring, certification renewal)
  - b. Temporal variability: historical data availability, frequency of updates, and possible lag times between data collection and availability
4. **Leverage Standard Identifiers:** standardized identifiers or codes enable data linking across systems. Coding schemes commonly utilized include:
  - a. Subject area codes: standardized codes for identifying courses or subject areas, and aligning to academic standards
  - b. School or district identifiers: unique identifiers assigned to schools and school districts (e.g., NCES codes)
  - c. Educator identifiers: unique, persistent educator IDs that are assigned when a teacher enters the workforce, and persist as they progress through the teacher pipeline
  - d. Program codes: common codes or nomenclature for identifying EPPs (e.g., CIP codes)
  - e. Geographic codes: for designating schools, districts, and regions

**5. Documentation:** to support users in understanding and utilizing the data dictionary. Documentation should provide clear instructions and continue to be maintained and supported in the following ways:

- a. Include “how to use” guidelines: provide the purpose and usage guidelines for the data dictionary, and how to maintain adherence with the data standard and use coding schemes correctly
- b. Implement version control: ensure strict versioning and change management processes are followed to make sure everyone is using the latest information
- c. Provide regular updates: establish procedures for periodic review and updates to reflect changing policy needs and data requirements
- d. Provide training materials: develop user-friendly guides and training resources to support data dictionary implementation
- e. Include example use cases: including examples and frequently asked questions can reduce friction when learning to first utilize the data dictionary

Before moving onto the next section, ask: “How can our data dictionary be improved, using some of the steps provided, to alleviate current challenges regarding standardized data definitions and reporting?”



- **[Developing State Data Specifications | Ed-Fi Alliance:](#)** this guide expresses to SEAs how to translate their existing data specifications according to the Ed-Fi Data Standard.
- **[Ed-Fi SEA Implementation Playbook:](#)** provides practical guidance for SEAs implementing Ed-Fi standards.
- **[Ed-Fi Data Standard:](#)** encompasses the full documentation and Ed-Fi Data Model handbooks for each version of the standard (updated annually).
- **[CEDS Data Elements:](#)** this website hosts standardized definitions, formats, and relationships for common education data elements.

## BRIGHT SPOT EXAMPLE



- The Kentucky [Data Dictionary](#) (format: XLSX, source site: [KY Center for Stats](#))
- The [RAPIDS data dictionary](#) (developed by the ED-Fi Alliance for Registered Teacher Apprenticeship Program data)
- The Texas Education Agency developed a data dictionary for a specific dataset pertaining to uncertified teachers across TX school systems: [Uncertified Teachers by School Systems 2024-2025](#) (format: XLSX, source site: [TEA Educator Reports and Data](#))

## 4.2 DEFINE KEY TERMS

**Objective:** Clarify complex concepts and ensure stakeholders share a common language and use a consistent set of definitions when referring to teacher workforce data indicators.

### Overview

Most common data elements are already addressed in the data standards (described in the prior section). States should avoid redefining standardized definitions. However, some terms are especially nuanced and locale or context specific, requiring states to curate their own standardized definitions and methods of measure. Furthermore, federal, state or other types of accreditation reports may have standardized definitions that differ from district or regional definitions. Even if your state's districts opt to use alternative definitions, it is good to be aware of standardized definitions, and encourage use of them whenever possible. This helps to ensure data holds the same meaning across districts and states, and gets everyone speaking a common language, enabling state-to-state comparisons.

When a standardized definition does not exist, or a state/districts opt to curate an alternate set of definitions, developing this common language should involve all key stakeholder groups. Coordination and achieving consensus among Data Providers, Data Stewards and Data Users may seem daunting, but agreeing on common definitions will improve the quality and usability of data long-term.

Below are a few examples of terms or concepts that we have witnessed as often being the most difficult to standardize:

- **Vacancy:** states often have differing definitions for what constitutes a teacher vacancy. Some base it off of strictly unfilled positions, while other states include positions filled by teachers who are not certified in the correct subject area, or positions filled by teachers with emergency or temporary certifications in their vacancy counts.
  - Recommendation: Establish a tiered definition system that distinguishes between "hard vacancies" (completely unfilled positions), "qualified vacancies" (filled by uncertified or emergency-certified teachers), and "assignment vacancies" (teachers teaching out of their certification area). This allows districts to report granular data while still rolling up to a common statewide metric. In addition, specify a census date (e.g., day 20 of the school year) for consistent measurement timing.
- **Shortage:** a shortage is often referred to as a school/district having reached a specified proportion of unfilled positions. Insufficient staffing is most often the way a shortage is defined, however, other ways to define it may be based on the number of qualified candidates available in comparison to the number of open job positions.
  - Recommendation: Differentiate between "current shortage" (unfilled positions at a given time) and "pipeline shortage" (insufficient qualified candidates in the labor market). Define shortage areas by subject and grade level using a consistent threshold (e.g., vacancy rate above X% or applicant-to-position ratio below Y). Align with federal shortage area designations where possible to facilitate comparison and grant eligibility.
- **Type of Educator Prep:** designating traditional from non-traditional educator prep programs (e.g., residency or apprenticeship programs) is sometimes unclear.
  - Recommendation: Create clear criteria based on program structure, for example, traditional programs might be defined as university-based with student teaching, while alternative programs include any pathway with simultaneous employment and coursework. Maintain a state-approved list of programs categorized by pathway type.
- **Retention and Attrition:** districts may have inconsistent timeframes and calculation methods for measuring teacher turnover.
  - Recommendation: Standardize on year-over-year retention measured from October to October (aligning with common reporting periods). Define separate metrics for "stayers" (same school), "movers" (different school in state), and "leavers" (exited profession or state). Specify exclusions (e.g., retirements, long-term leaves) and require consistent denominator calculations to enable valid comparisons across districts.

- **Teacher Quality:** measures of teacher quality or teacher effectiveness are particularly tricky, and there are many factors that could be used to determine whether a teacher is deemed effective or not. For a few examples, refer to the [Educator Equity Best Practices Guide](#) NCTQ has developed.
  - Recommendation: Rather than attempting one composite measure, establish a framework with multiple standardized indicators: credentials (certification, degree level), experience (years teaching), evaluation ratings (using the state's approved framework), and student outcomes (value-added or growth measures where available). Allow districts to report on all indicators while being transparent that "quality" is multidimensional. Consider which indicators will be used for specific policy decisions and communicate those clearly.

Finally, the most important aspect is to come to a consensus on how your state, districts, and individual schools define these terms that are directly related to how teacher shortages are routinely identified and assessed. It's critical to make sure these definitions are clearly documented, communicated and understood among stakeholders.

# COMPONENT 5: DATA INFRASTRUCTURE IMPLEMENTATION

TECHNICAL

—How does one build a data system? Let's get technical!

States must tackle the complex technical work of actually building a data infrastructure that will power their teacher workforce insights. This component is the most intensive phase of system development, requiring careful architectural decisions, robust security implementations, and skilled technical execution. The infrastructure should be designed to handle diverse data formats, support real-time and batch processing of data, provide secure data sharing, and ensure reliable performance as the system scales. States face critical decisions on whether to host their data system on the cloud or on-premises systems, what database architecture will support the disparate datasets that will be collected, where integration with external systems may be beneficial, and what security controls should be built into the system to streamline compliance and data protections from the ground up. The technical choices made during this phase will determine the system's long-term scalability, maintainability, and flexibility to evolve with changing needs. Success requires not only solid technical expertise but also agile project management and close coordination between the technical and data governance teams. Given the complexity and specialized nature of this work, states must carefully evaluate their internal capacity and consider partnerships with MSPs or technical consultants to ensure successful implementation.

## 5.1 DESIGN AND BUILD THE ARCHITECTURE

**Objective:** Create a comprehensive technical blueprint that defines the system architecture, infrastructure requirements, and specifications to guide the implementation of the teacher workforce data system.

## Overview:

System architecture design is the foundation of successful data infrastructure implementation. This phase requires translating business requirements into technical specifications—requiring important decisions to be made around cloud vs on-prem hosting, database architectures, processing capabilities, and integration approaches. The technical team is charged with informing an optimal approach for system hosting—exploring cloud-based solutions versus locally-hosted systems—and database providers, considering factors such as cost, scalability, security, and maintenance requirements.

## Key Architecture Decisions:

### 1. Hosting Infrastructure:

- a. Cloud-based Systems: leverage cloud platforms (e.g., AWS, Microsoft Azure, Google Cloud Platform) for compute and data storage
  - i. Benefits: reduced infrastructure management, automatic scalability, built-in security features, and disaster recovery capabilities
  - ii. Considerations: ongoing costs increase with use, potential for vendor lock-in
- b. On-premises Systems: deploy locally-hosted compute and data storage infrastructure
  - i. Benefits: direct control over data and systems, potential cost savings for large implementations
  - ii. Considerations: higher upfront cost (e.g., to buy servers) but few to no recurring usage costs, responsible for ongoing maintenance and security, limited scalability

### 2. Database Structure:

- a. SQL/Relational Databases (e.g., PostgreSQL, MySQL, Amazon RDS):
  - i. Best for: structured data with clear relationships, complex queries, strong consistency requirements
- b. NoSQL/Flexible Databases (e.g., MongoDB, DynamoDB):
  - i. Best for: semi-structured or unstructured data, rapid development, horizontal scaling needs
- c. Data Lake (e.g., Amazon S3, Databricks): repository for storing raw data in its native format until needed for analysis
  - i. Best for: storing diverse, large and raw volumes of structured or unstructured data (e.g., PDF certification documents, CSV exports, student information system backups)

### 3. Integration and API Design:

- a. API Endpoints: develop standardized interfaces and APIs for data submission and retrieval. Types of APIs include:
  - i. RESTful APIs for simple data operations (e.g., create, read, update, delete)
  - ii. GraphQL APIs for complex data queries and automatic validation
- b. Data Integration Capabilities:
  - i. Provision data pipelines to automate routine data imports from Data Provider systems (e.g., HR systems, student information systems, and preparation programs)
  - ii. ETL (Extract, Transform, Load) processes for data cleaning and standardization
  - iii. Real-time streaming capabilities for time-sensitive data updates

## BRIGHT SPOT EXAMPLE



The Ed-Fi Alliance is a community-driven, open-source suite of tools to support state education agencies, technology providers, collaboratives, and other organizations in adopting a common language and associated technical implementations enabling education technology systems to securely exchange data from multiple sources. The following are tools provided via the Ed-Fi Technology Suite:

- Ed-Fi Data Standard
- Ed-Fi Unifying Data Model
- Ed-Fi ODS/API Platform

The **Ed-Fi Data Standard** refers to the overall set of data standards developed by Ed-Fi to facilitate data interoperability in education systems. It is a framework for how educational data should be structured, exchanged, and managed across systems in a way that allows different software and platforms to communicate with each other. The data standard focuses on defining common formats, data elements, and protocols for things like student information, assessments, courses, and staff.

The **Ed-Fi Unifying Data Model** is part of the Ed-Fi Data Standard and serves as the specific model or schema for structuring educational data within that standard. It provides a detailed and comprehensive way of organizing and representing educational data. This includes how data entities (like students, teachers, and classes) relate to each other and the attributes those entities should have. Essentially, it is a reference model that is used to design data systems, ensuring consistency and interoperability when data are exchanged between systems.

The [Ed-Fi ODS / API Platform](#) is a concrete implementation of a relational database and a companion API harmonized with the Ed-Fi Data Standard. The operational data store (ODS) implementation uses Microsoft SQL Server or PostgreSQL as a database platform. The Ed-Fi Unifying Data Model (UDM) provides the basis for the data transferred and manipulated by an Ed-Fi REST API implementation.

The following are links to Ed-Fi implementation guides and technical reference documents:

- [Ed-Fi Data Exchange Standard](#)
- [Ed-Fi API Design and Implementation Guidelines](#)
- [Understanding UML, REST API, and Database Expressions of the Ed-Fi Data Model](#)
- [Ed-Fi ODS / API for Suite 3 v7.3](#)
- [Ed-Fi Implementation Guide for State Education Agencies](#)

## 5.2 SPECIFY DATA SECURITY AND PRIVACY REQUIREMENTS

**Objective:** Establish comprehensive security controls and privacy protections that safeguard sensitive teacher workforce data while enabling authorized access and use.

### Overview

Given the sensitive nature of teacher workforce data, including personally identifiable information (PII) and employment records, robust security and privacy measures are essential. The system must comply with federal privacy laws, state regulations, and industry best practices while maintaining usability for authorized users.

*Note: Rich disaggregated data (e.g., by location, subject area, or certification type) is needed to understand the strengths and weaknesses of the teacher pipeline across all types of educators. Furthermore, only a handful of states collect and make accessible comprehensive and disaggregated forms of data to the states' major stakeholders, due to legitimate data privacy and security restrictions. However, with careful data de-identification protocols and sophisticated access controls, high-quality, dis*

## Security Control Implementation:

### 1. Access Control Systems:

- a. Role-Based Access Control (RBAC): implement granular permissions based on user roles and responsibilities
  - i. Define user roles (e.g., system administrators, data analysts, district users, researchers) that directly correspond with the level of access to data
  - ii. Assign specific permissions for data access, modification, and sharing capabilities (e.g., read, write, delete)
  - iii. Implement principle of least privilege required to minimize access to data not required to perform normal business functions
- b. Multi-Factor Authentication (MFA): require multi-step verification for system access, especially for users with administrator access
- c. Single Sign-On (SSO): integrate authentication processes across existing state systems to provide a single sign-on procedure

### 2. Data Protection Measures:

- a. Data Encryption in Transit: secure all data transmission using TLS/SSL protocols
- b. Data Encryption at Rest: encrypt stored data using industry-standard encryption algorithms
- c. Key Management: implement secure key storage and rotation procedures

### 3. System Security Controls:

- a. Network Security: implement firewalls, intrusion detection systems, and network segmentation
- b. Audit Logging: maintain comprehensive logs of all system access and data operations
- c. Incident Response: develop procedures for detecting, responding to, and recovering from security attacks or data breaches
- d. Data Backups: implement automated backup procedures with tested recovery processes
- e. Regular Security Audits and Updates: conduct periodic vulnerability assessments, and maintain current security patches and software updates



- [Open Source Guidebook on Advanced R&D in Education](#): a practical guide to building the future of education with R&D
- [SETA-ED](#) offers critical supports in implementing and maintaining data systems, including data privacy and security services.
- [NIST Cybersecurity Framework](#): a robust framework for implementing

## Data Privacy Requirements:

### 1. Regulatory Privacy Compliance:

- a. FERPA (Family Educational Rights and Privacy Act): requires the following controls for ensuring the protection of education records
  - i. Define legitimate educational interests for access to data
  - ii. Establish procedures for sharing data with external parties
  - iii. Implement audit trails for data access and disclosure
- b. State Privacy Laws: systems will need to comply with state-specific privacy regulations and requirements (e.g., regional data storage, 3rd-party security audits, data deletion, and right to be forgotten protocols)
- c. District Data Sharing Agreements (DSAs): the Data Governance team is responsible for ensuring privacy protections specified in DSAs are implemented and adhered to through the management and use of data

### 2. Principles for Implementing *Privacy-by-Design*:

- a. Data Minimization: collect and retain only data necessary for specific purposes
- b. Purpose Limitation: use data only for stated purposes and obtain consent for new uses
- c. Transparency: provide clear information about data collection, use, and sharing practices
- d. Individual Rights: implement procedures for data subject access requests and corrections
- e. Data Retention: establish clear data retention schedules and secure deletion procedures

## BRIGHT SPOT EXAMPLE



**Utah's Office of Data Privacy** is a bright spot for providing assistance from the state to various state agencies in building solid privacy programs around the data that those agencies collect and manage. The Office provides a roadmap for implementing a privacy program, virtual trainings and workshops, and various templates for privacy policies, data processing agreements (DPAs), and more. Guided by Utah Code and Board Rule, the Utah State

Board of Education also has its own **Student Data Privacy division**. This division provides data privacy information for LEA Data Managers, Information Security Officers, Educators, and Parents. The division provides online courses and templates to assist LEA Data Managers in forming DPAs with other vendors and agencies, and even a **YouTube channel** to explain federal laws and data privacy and security basics. These resources are not specific to a teacher workforce data system, but they are a great starting point and demonstrate that the state as a whole prioritizes upholding data privacy across all its data-driven endeavors.

# COMPONENT 6: DATA COLLECTION MECHANISMS AND INTERVALS

TECHNICAL

—Make data submissions as simple and easy as possible

Effective and useful teacher workforce data systems require well-designed and simple interfaces for streamlining frequent data collection from various Data Providers. Easy-to-use mechanisms for submitting data not only reduce stress and burden on Data Providers, but also help to ensure consistent data quality in each round of data collection, including uniform file formats (e.g., CSV, JSON, XML), naming conventions, and utilizing a common data standard.

## 6.1 DEVELOP CLEAR AND EASY DATA COLLECTION PROTOCOLS

**Objective:** Work with each Data Provider and the technical team to develop mechanisms for streamlining data collection.

### **Data Collection Mechanisms:**

The data system should be capable of leveraging different mechanisms for collecting data, in order to make it the most efficient and compatible with the various source systems (i.e., the systems that the data are coming from). Some possible options for integrated data collection mechanisms include:

1. Manual Entry Tools with Standardized Templates: Data Providers fill in their data using clear templates and a web-user interface.
  - a. Benefits: good for Data Providers with limited technical capacity, and ensures business rules and standards for data are adhered to upon data submissions
  - b. Considerations: this option is the most time-intensive for Data Providers, and can be challenging or not feasible for Data Providers to make updates or corrections to past data submissions

2. Bulk Upload Portals: secure portals for batch uploads of data files
  - a. Benefits: easy for Data Providers to quickly upload and share data
  - b. Considerations: typically unstructured data with little to no data standardization
3. APIs for Automated Data Collection: enable APIs for automated or on-demand system-to-system transfers of data
  - a. Benefits: excellent for real-time or high-frequency data exchange, built using a data standard by design, and once in place, automated APIs can dramatically reduce the amount of human labor required to execute data submissions
  - b. Considerations: least time-intensive option for Data Providers, requires a well-defined data standard to be in place, tends to have higher up-front costs to set up, and can be difficult to navigate security boundaries when integrating systems

## 6.2 DETERMINE DATA COLLECTION INTERVALS AND FORMAT

**Objective:** Determine what data are needed when, how often, and in what format.

### Data Collection Intervals:

It's critical to establish collection intervals that align with the natural rhythms of the education sector while providing timely insights for decision-making. Different types of workforce data require different collection frequencies: some information, like annual staffing snapshots, follow predictable academic calendar patterns, while other data elements, such as real-time vacancy tracking, demand more frequent updates to support responsive workforce planning. The following are some practices for data collection intervals:

- Annual collections:
  - Commonly used for: EPP data (e.g., enrollment data, completion rates, placement data)
- Quarterly or Semester collections:
  - Commonly used for: employment data snapshots and educator retention/attrition information (e.g., workforce exits and transfers)
- Near Real-time collections:
  - Commonly used for: vacancy data (e.g., open positions, late hires, substitute fill rates) or scraping of job board postings

### Specify Data Formats:

Streamlined or automated data collection solutions not only help to ease routine data collection needs, but they also can ensure consistent structure and standardization of data. Across each of these options, you can build in rules to validate that data are in the proper form at the point of entry. The following are examples of some rules that you can embed into your APIs and standardized templates:

1. Data types: specify the appropriate data type (e.g., text, numeric, date, boolean) for each data element
2. Field lengths: define maximum character limits and formatting requirements
3. Required vs. optional fields: designate which data elements are mandatory versus optional for different data collection activities
  - a. Validation rules: by using technical solutions for data collection, it is possible to define custom validation rules (e.g., acceptable value ranges or categories) to automatically check each data value for compliance and proper format at the point of entry—for more details see section 7.1: Data Validation and Quality Assurance Processes

## BRIGHT SPOT EXAMPLE



**Washington State's Education Research and Data Center (ERDC)** stands out as a national bright spot for teacher workforce data systems due to their **clear and standardized guidance for data submissions**. ERDC provides comprehensive resources, including pre-formatted Excel templates with proper data typing for each collection, lookup tables, and diagnostic guidelines for data cleanup prior to submission. Their approach includes clear due

dates, comprehensive FAQs addressing common submission concerns, and detailed documentation that walks Data Providers through each step of the process. This systematic, user-centered approach to data collection guidance significantly reduces the administrative burden on districts and EPPs while ensuring high-quality, consistent data submissions that enable Washington to maintain robust workforce analytics and respond effectively to teacher supply and demand fluctuations across the state.

# COMPONENT 7: CLEAN, QUALITY CHECK, AND MERGE DATA

TECHNICAL

—Join data together to enable longitudinal and cross-sector data analysis

Siloed data across agencies and institutions is a common challenge to collectively analyze workforce trends. To effectively track teacher supply and demand over time and understand how factors along the teacher pipeline influence these dynamics, states need integrated data sources. This integration allows for comprehensive longitudinal and intersectional analyses.

## How do we get different sources of educator data to talk to one another?

A crucial first step is validating that the collected data conforms enough to a common standard, making it possible to join together. When data doesn't follow a consistent schema or share in how they measure and define key elements, it's almost impossible to integrate those disparate data sources together. Following the steps in **Component 4: Standardizing Data**, should ensure data are ready to join with other datasets.

## How does joining data across sources help us measure and respond to teacher supply and demand?

[The Ed-Fi Playbook: Advancing Data Utilization to Support Educator Preparation and Teacher Workforce](#), explains that states should focus on building data systems that make key teacher pipeline information—such as applications, enrollments, completions, certifications, and employment—readily available. They should also track retention, mobility, credential alignment, and teacher shortages. By linking these datasets to information on vacancies, certifications, and turnover, states can generate timely, accurate insights into the teacher workforce and more effectively address imbalances between supply and demand.

# 7.1 DATA VALIDATION AND QUALITY ASSURANCE PROCESSES

**Objective:** Implement systematic mechanisms to ensure data are valid, accurate, and trustworthy.

## Data Validation:

Collecting insightful and usable teacher workforce data over time depends on having standardized, consistent, and high-quality data. Data validation and quality assurance processes ensure data are accurate, complete, and consistent across multiple data collection cycles. While it requires being proactive and designing a robust system of validation rules and checkpoints from the onset, systematic validation reduces human-intensive manual review and validation processes. Once implemented, systematic validation can free up human capacity to focus on gathering data insights and workforce planning efforts.

Systematic validation checks to make sure that submitted data adheres to the expected structure and follows certain rules, for instance:

- The values adhere to the proper data type (e.g., text, numeric, date, boolean) for each data element.
- The values are logical according to expected ranges (e.g., reported class sizes are within reasonable bounds, or salary figures align with state pay scales).
- There are no inconsistencies (e.g., a teacher's years of experience decreases from one reporting period to the next, or where demographic information conflicts with previously submitted records).
- Pre-specified logic/rules flag common errors or special cases (e.g., a teacher's reported subject area doesn't match their certification).
- Missing data are identified early (e.g., at the point of data collection).

Systematic validation protocols can also provide immediate feedback to the Data Provider as to whether they need to modify or correct any of their data prior to submission—reducing the rate of potential data fixes needed post-submission.

## Data Quality Checks:

While automated systems can catch many errors early and quickly, manual reviews contribute contextual understanding and domain expertise essential for ensuring quality data. These audits should focus on most error-prone or critical areas of data, and include systematic sampling across data submissions to check for data consistency and quality. Manual review processes allow trained staff to evaluate nuanced issues that automated systems might miss, such as proper coding of unique employment arrangements or accurate classification of teacher roles that span multiple subject areas. Human reviews also provide valuable insights into common data submission challenges that can inform improvements to collection processes and guidance materials.



### KEY RESOURCES

**[Embracing Data Validation with the Ed-Fi ODS/API](#)**—this resource assumes you have implemented the Ed-Fi ODS/API, and discusses the data validation steps that can be automatically executed upon sending data to the API or once it is stored in the ODS. In addition, this guide offers rationale for why proactive data validation unlocks success, tips for defining data validation rules, and workbooks and templates to follow along on your own.



### BRIGHT SPOT EXAMPLE

**[Wisconsin's WISEdata Validations API](#)** stands out as a bright spot for systematic data validation. Wisconsin's approach appears to go beyond basic formatting checks to a multi-layered validation system that includes Layer 1 (L1) validations through the Ed-Fi API and enhanced Layer 2 (L2) validations through a downstream validation engine that applies specific business rules. What makes Wisconsin particularly impressive is a systematic approach to

validation via an API. Their website describes detailed validation rules with unique identifiers, clear descriptions, help links to knowledge base articles, severity levels ranging from warnings to major validation errors, and categorization by domain areas like enrollment, special education, and budget. Wisconsin's system claims real-time validation results that Data Providers can consume directly through the API, creating immediate feedback loops that help Data Providers identify and correct issues efficiently. This systematic approach transforms data validation from a black-box process into a transparent, collaborative system where districts receive actionable feedback with specific error codes, context, and guidance for resolution, ultimately improving data quality.

## 7.2 LINK AND JOIN DATA FROM MULTIPLE SYSTEMS

**Objective:** Identify common variables to link across various sources of data and join those together for comprehensive longitudinal and integrative analyses.

**Overview:**

Creating a comprehensive view of teacher workforce dynamics requires the ability to seamlessly connect data across data sources and timeframes. Effective data linking unlocks longitudinal analyses—observing individuals throughout the entire teacher pipeline, from preparation through career—by transforming isolated data into consolidated datasets.

**Extract, Transform, and Load Data:**

Pulling together disparate data requires data to be *extracted* from the various source systems, *transformed* to comply with a consistent format and data standard, and *loaded* into a central database—known as ETL pipelines. These mechanisms might be the same across the various datasets or slightly different to accommodate different Data Provider needs and data structures. Extracting data occurs via the data collection process (e.g., user interface, API, bulk uploads). Once collected, data are transformed and validated to ensure it follows a consistent data standard and any pre-specified business rules. Once in the proper format, data are loaded into a central database or data store, where it can then be joined with other data tables and accessed for downstream analyses.

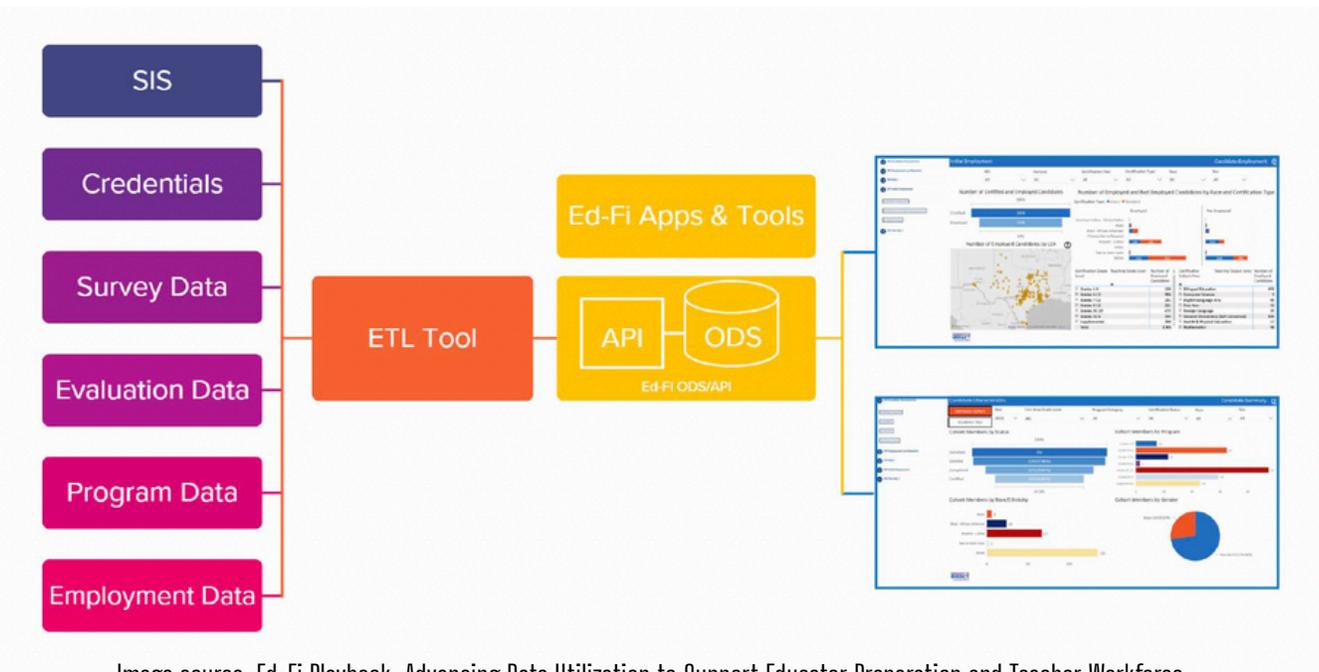


Image source: Ed-Fi Playbook: Advancing Data Utilization to Support Educator Preparation and Teacher Workforce

### **Establishing Common Identifiers Across Datasets:**

The foundation of successful data joining lies in implementing identifiers or common codes that can reliably connect records across diverse data sources. SEAs should establish (or use existing) identifiers that persist across systems and time. We advise using identifiers that are non-sensitive (e.g., certification numbers, randomly generated teacher IDs) instead of personal identifiable information (e.g., SSNs), when possible, to protect teachers' privacy and information.

Maintaining a consistent identifier across the teacher pipeline may not always be possible. The identifier strategy must account for the reality that different systems may use varying formats or coding schemes. For example, EPPs might use institutional student IDs, certification boards may assign professional license numbers, and school districts often maintain their own employee IDs. Creating crosswalk tables across the different identification systems is a potential workaround to this challenge. However, establishing common identifiers across disparate systems is the best way to ensure that data can be easily joined.

### **Tracking Professional Trajectories Longitudinally:**

Longitudinal data linking enables states to follow cohorts of teachers through career transitions—from preparation and certification to placement and retention. Longitudinal data can unlock powerful insights into workforce patterns and outcomes. For example, connecting EPP data with initial certification and placement records allows states to analyze different preparation pathways and identify factors that influence successful program completion, where those teachers are placed in the workforce. As teachers progress through their careers, linked data can help states track movements between districts, schools, and positions, providing insights into career mobility patterns and factors that influence retention. This longitudinal view becomes particularly valuable for understanding teacher workforce dynamics, including temporary exits and re-entries, transitions into leadership roles, and movements across regions/districts. Successful longitudinal data linking depends on consistent data collection practices that maintain continuity across time periods and organizational changes. See **Component 4: Standardizing Data** and **Component 6: Data Collection Mechanisms and Intervals** for further guidance on data collection protocols, and specifying required fields, data types, and reporting timelines across Data Providers.



## Data Privacy and Security Spotlight

### Linked Data Exposes New Risks:

Linked or longitudinal datasets raise additional privacy and security risks, including: identity re-identification, consent violations, group discrimination, and eroding trust. Historical data that seemed harmless when collected can become sensitive or potentially identifiable when linked with newer information. These potential risks must be addressed through intentional data governance and information security. The data governance team should establish clear policies for what data are allowed to be shared and used for research purposes. Automated controls that can help ensure compliance include: access controls and least privilege policies, data de-identification protocols, and using synthetic identifiers for research purposes while maintaining the ability to re-identify records when necessary, only for operational purposes. Additionally, states should establish clear data retention policies that specify how long linked data will be maintained and under what circumstances records may be purged or archived.

### BRIGHT SPOT EXAMPLE



Arkansas has been able to link Title II data with licenses, placement, and even retention. Arkansas analyzes data from the various systems by linking to educator identifiers. They integrate data across the following systems: Statewide information system (SIS) for district data reporting and LEA insights, the Teacher Excellence and Support System (TESS) for educator evaluation data, state licensure system, and statewide educator surveys. The

The following report provides some examples of the analyses they have been able to do using this longitudinal data, particularly around retention: [\*\*2024–25 Arkansas Teacher Retention: Statewide Stability Amid Ongoing Local Challenges\*\*](#). Shortages are calculated and categorized according to course codes and cross-walked with licensure codes to identify current and future teacher supply gaps.

Tennessee’s annually published [\*\*Educator Preparation Report Card\*\*](#) lists EPPs across the state and data on those programs mandated by the state’s legislature. The data provided includes placement rates (the percentage of program completers who are hired in Tennessee public schools), one-, two-, and three-year retention, program outcomes, and teachers’ reported satisfaction rates. These dashboards provide insights into how well the state’s EPPs are supplying teachers across the state, and how they fare in the job market and the classroom.

# COMPONENT 8: DATA VISUALIZATION & REPORTING

TECHNICAL

—Promote equity and enhance the value of your data simply by making it accessible

An important, yet often missing, capability is the ability for non-technical Data Users to access data of their own accord. Oftentimes, states have only 1-2 people who are intimately familiar with the data stores and can pull requested datasets. Without institutional knowledge, navigating complex data stores tends to be difficult and burdensome, and can sometimes lead Data Users to incorrectly assume that the data they need doesn't exist. [A report from NCTQ](#) highlights states that were found to be reporting teacher supply and demand data in a way that is publicly accessible for stakeholders to use. The report aims to convince stakeholders of how state-level data reporting can highlight teacher shortage challenges and rally support and resources across the state.

In order for a wide array of Data Users to be able to understand, access, and leverage data, it needs to be easily accessible and relatively simple to find and pull the data they are interested in, without violating any data security or privacy protections. This component provides some ideas around how data can be shared and made accessible, with appropriate access controls still in place to protect sensitive data. The solutions include dynamic dashboards, data visualizations, automated static reporting, and even AI-powered data exploration.

## 8.1 DYNAMIC DASHBOARDS AND DATA VISUALS

**Objective:** Display data in dynamic, interactive, and user-friendly formats (e.g., data dashboards and visualizations) to meet the needs of the Data Users and stakeholders.

**Dynamic Data Dashboards** can be used to display scheduled collections or live streams of data. Most dashboards are dynamic—meaning that as data is updated or added to the backend data system, the data reported on the frontend dashboard gets updated automatically. Data dashboards are not only great for exploring what data are currently available, but they're also helpful for monitoring ongoing data submissions. For example, dashboards can report how much data has been collected in the cycle so far, whether there is missing data (from specific Data Providers, or more broadly), whether there are missing categories or elements of data, and whether the data has successfully passed or failed any data validation rules. These types of dashboards are particularly useful for helping Data Stewards monitor and manage data collections, by automatically checking that submitted data are complete and pass systematic data validation checks.

*Key features of optimal data dashboards: automatically update with new data, show high-level information on what data currently exists, and provide insights on the quality and validity of submitted data*

**Interactive Data Visuals** offer Data Users the ability to interact with data, without putting raw data or sensitive educator information at risk. Static data visuals display data in graphs or charts to visualize patterns and trends, but static visuals are snapshots of data and don't allow users to interact with the data or explore areas of data more deeply. Interactive data visuals offer controls to the user (e.g., ability to filter and hone in on specific data elements) for more hands-on exploration of the data. Options to filter data and adjust how data are displayed enable Data Users to get the right level of information they need without needing to coordinate with the Data Steward or analyst—dramatically speeding up data access and exploration. Interactive data visuals are a great way to enhance the accessibility of the data for different end-users (e.g., district managers, EPPs, policymakers). Softwares such as Tableau, Power BI, or Looker Studio are common interactive data visualization toolkits.

*Key features of interactive data visuals: user interfaces are intuitive and easy to use, and provide the ability for Data Users to filter, disaggregate, and manipulate data to explore and analyze specific contexts.*

## BRIGHT SPOT EXAMPLE



Illinois Teacher Shortage Dashboard ([2023-2024 Educator Shortage – IARSS](#)) is hosted using Power BI and is developed and updated each year by a technical consultant to the state. The data powering this dashboard was collected by the Illinois Workforce and Education Research Collaborative (IWERC) to provide context on how education leaders discover qualified candidates, causes for open positions, methods schools use to accommodate unfilled

positions, and the impacts of current solutions being implemented or proposed. Users can navigate to specific data that is applicable to them and leverage the data filters to further explore the data. The interactive data visuals support a wide net of end-users (from district leaders to policymakers) in accessing usable information to identify issues pertaining to teacher shortages in Illinois.

Colorado [Educator Shortage Dashboard](#) describes shortages by district, subject area and trends over time.

Massachusetts [DESE Directory of Datasets and Reports](#) publishes an [Educator Data Dashboard](#)—displaying state and district-level educator data by race and ethnicity. This dashboard focuses mainly on the state of teacher supply, including data on currently employed teachers and teachers in the pipeline.

## KEY RESOURCE



[Ed-Fi Alliance Open-Source Educator Workforce Use Case Tool](#) — This resource presents a structured approach to addressing teacher shortages through data-driven dashboards and visual reporting tools. The Ed-Fi Alliance created a set of prompting questions and visual design protocols for building interactive data visuals aimed at addressing critical staffing challenges. Also provided is a list of the required data elements needed to answer those

questions—mapped to the Ed-Fi Standard—to help agencies scope their data needs.



## Data Privacy and Security Spotlight

**Tiered Access Controls:** Implementing appropriate access controls is critical for maintaining data privacy while maximizing the utility of workforce information. States should first separate data according to what is allowed to be public versus what needs to remain internal, and then get more granular as needed on who should access internal data assets and possess certain types of functionality—via role-based access controls. Tiered levels of access help states accommodate different types of Data Users, while ensuring compliance with requirements to protect data. The following are three tiers that serve as helpful starting points for delineating access to data:

- **Publicly available data assets:** include interactive data visuals and reports available on the public web. Public data tends to be aggregated to reduce risks of re-identification, while still providing valuable insights into workforce trends and system-wide patterns. Public interfaces often provide data on overall teacher supply and demand, certification area shortages, geographic distribution patterns, and retention rates at appropriate levels of aggregation.
  - Note: a general rule of thumb in data privacy is to establish a ***k-anonymity*** threshold. The core idea is that each individual's data within the dataset should be indistinguishable from at least  $k-1$  other individuals based on indirect identifying factors in the data. This means that for any combination of quasi-identifiers (e.g., school, grade-level, race/ethnicity), there are at least  $k$  individuals with the same values, making it difficult to link a data instance back to a specific individual. The value of  $k$  determines the level of anonymity. A higher  $k$  value ensures a greater level of privacy but can reduce the specificity of the data, and in some cases, the data's utility. An optimal level of  $k$  depends on the specific dataset, its sensitivity, and an acceptable level of risk of re-identification.
- **Internal data assets:** typically include data dashboards and backend de-identified datasets used to populate front-end interactive data visuals. This data is most often leveraged by the Data Stewards and data analysts with specific access to the data. Data Providers (e.g., district-level users) may access these assets to ensure the data they submit is accurate and complete, but they should only be able to access data relevant to their jurisdiction, while Data Stewards users may have broader access for ensuring data quality and overseeing use of the system.
- **Sensitive data assets:** include educator PII or school-identified data—these data should be managed with restricted access via granular permission systems, enabling Data Stewards to ensure all access to sensitive data complies with applicable data sharing agreements and privacy regulations.

## 8.2 AUTOMATED STATIC REPORTING

**Objective:** Streamline routine data reporting requirements through automated data pipelines.

**Common types of reports:**

- Annual Reports: print and web-based reports with year-over-year trends (e.g., “State of the Educator Workforce” report, or federal data reports)
- One-Pagers and Briefs: for legislators, regional briefs for district planning, or general updates for public consumption

**Ways to automate:**

- Standardize the method for querying data: if the same data is queried on a routine basis, the API request or SQL query should be saved to make sure data is pulled efficiently and consistently each time.
- Configure data pulls to occur on time-based or event-based intervals: if the same data needs to be pulled on a recurring basis, configure a job to pull that data at specific points in time or following specific events (e.g., data submissions or validation checks).
- Automate ETL (Extract, Transform, Load) pipelines: implement automated jobs on timed schedules to extract the data from the data source, transform the data into the needed form, and load it into the target system or report.

BRIGHT  
SPOT  
EXAMPLE



Louisiana's Department of Education [annual data reports](#) encompass teacher evaluations, credentials, salary information, headcounts, exit surveys and overall workforce reports. The educator workforce reports provide an overview of workforce data for over 1,300 public schools and charter schools in Louisiana.



## Emerging NEW Technology: LLM-Powered Data Exploration

Traditional workforce data systems often create a bottleneck where only technical staff with database expertise can explore and export data from the system. This limitation restricts interacting with data to a small group of technical and data-savvy people, who understand complex query languages, database schemas, and data transformation processes.

### The Solution:

Integrating artificial intelligence (AI), via large language models (LLMs), into teacher workforce data systems can democratize data access by allowing stakeholders to query databases using natural language. Instead of requiring database knowledge and technical expertise, users can ask questions like:

- "Show me teacher retention rates by subject area for the past three years."
- "Which districts have the highest vacancy rates in special education?"
- "How many newly hired teachers have alternative certification by region?"

**How It Works**—The LLM acts as an intelligent intermediary that:

1. Interprets natural language questions from users
2. Translates those questions into SQL queries or API requests
3. Executes the queries against the workforce database
4. Returns results in user-friendly formats, such as natural language answers or exportable data tables

### Benefits for Workforce Data Systems:

- **Ease of Access:** various stakeholders can directly explore data without technical intermediaries.
- **Faster Insights:** real-time querying eliminates delays waiting for data requests to be processed.
- **Exploratory Analysis:** users can follow up with additional questions, enabling rapid and deeper investigation of workforce trends.
- **Reduced Technical Barriers:** non-technical stakeholders can participate more fully in data exploration and analysis.
- **Built-in Security and Privacy Controls:** developers can encode guardrails (e.g., only returns de-identified or aggregate-level data) into AI systems to prevent leakage of identifiable information.

This technology transforms workforce data systems from static reporting tools into interactive, accessible platforms that empower all education stakeholders to extract insights that can inform policy development and strategic planning.

# COMPONENT 9: SUPPORTING DATA-DRIVEN INITIATIVES

OPERATIONAL

—Bridge data into action using data storytelling

Data-centric initiatives transform teacher workforce data into actionable intelligence that drives meaningful policy reform and addresses persistent educator shortages. States with robust workforce data systems can move beyond reactive emergency policies to proactive, evidence-based solutions that target specific challenges with precision and measurable impact. To do this, states should combine their robust data systems with impactful analytics and communications through *data storytelling*.

## 9.1 WHAT IS DATA STORYTELLING?

Data storytelling is the practice of combining data analysis with narrative techniques to communicate insights in a way that is both informative and compelling. It transforms raw numbers and statistics into meaningful stories that resonate with audiences, drive understanding, and inspire action. Rather than simply presenting data, data storytelling creates a coherent narrative that explains what the data means, why it matters, and what should be done about it.

At its core, data storytelling bridges the gap between technical analysis and human understanding. It takes complex datasets and makes them accessible, memorable, and actionable for diverse audiences who may not have technical expertise but need to make decisions based on the insights. Data storytelling can help others to see the value in robust workforce data systems by understanding the outcomes they bring. Furthermore, effective data storytelling builds trust amongst stakeholders and community members by being transparent about methodology, acknowledging limitations, and presenting findings in an honest, accessible way. When audiences understand how conclusions were reached and feel confident in the analysis, they're more likely to trust and act on the recommendations.

It's important to start with a clear, compelling question that your audience cares about. The question should be specific, answerable from the data, and of importance to your audience. For example, instead of "What does our teacher retention data show?" ask "Why are new teachers leaving our highest-need schools, and what can we do to keep them?" The following section provides examples of policy initiatives that could be used to develop informative questions.

## 9.2 IDENTIFY AND SUPPORT DATA-DRIVEN POLICY INITIATIVES

### Examples of Potential High-Impact Policy Areas and Data-Driven Approaches:

- Building support for teacher incentives and benefit programs:
  - Use retention data to identify which incentives are most effective for different teacher demographics and career stages;
  - Target professional development investments based on areas where teachers report the greatest need for support;
  - Design compensation packages that address the specific factors contributing to turnover in high-need areas.
- Prioritizing funding and support for high-need LEAs:
  - Utilize comprehensive shortage data to create criteria for identifying districts requiring additional support;
  - Track the effectiveness of interventions over time to refine funding formulas and support strategies;
  - Ensure resources reach the schools and subjects with the most acute needs.
- Creating and supporting non-traditional EPPs in shortage areas to help bolster teacher supply:
  - Identify geographic and subject-specific gaps that might benefit from alternative or creative forms of teacher preparation;
  - Use employment outcome data to evaluate the effectiveness of alternative certification programs;
  - Develop and support alternative teacher preparation pathways to increase supply in needed areas.
- Amending licensure/certification requirements or processes based on teacher supply and demand:
  - Analyze the relationship between preparation requirements and teacher effectiveness to inform policy changes;
  - Use data to identify barriers in certification processes that may be deterring qualified candidates;
  - Track outcomes of policy changes to ensure reforms achieve intended goals without unintended consequences.

**BRIGHT  
SPOT  
EXAMPLE**



**Illinois Teacher Vacancy Grant Program** exemplifies how states can leverage comprehensive workforce data to design, implement, and evaluate targeted policy interventions. The Teacher Vacancy Grant is part of a pilot program aimed at addressing chronic shortages by providing the state's most understaffed districts with resources to attract, hire, support, and retain teachers, allowing districts maximum flexibility to use allocated funds

in innovative, creative, and evidence-based ways. The state partnered with the Illinois Workforce and Education Research Collaborative (IWERC) to conduct ongoing evaluation of the program's effectiveness. Illinois used comprehensive vacancy data to identify the districts with the greatest need, ensuring resources reached the areas where they could have the most impact. In the program's first year, the state claims more than 51% of districts receiving grant funds saw a decrease in unfilled positions, compared to a smaller percentage (17%) of districts that saw a decrease but did not receive the grant.

An example of a data-driven impact story — **How Texas Is Setting Teachers – and Students – Up for Success**. This impact story helps to convey how timely access to key data can drive change and strengthen how EPPs prepare future teachers for success in the classroom. This story highlights how access to key data can move beyond simply fulfilling compliance requirements to spurring continuous improvements across EPPs and the state education system as a whole.

# COMPONENT 10: SUSTAINABILITY AND EXPANSION

OPERATIONAL

—An engaged user base lays the foundation for long-term impact and sustainability

The long-term success of teacher workforce data systems depends not on their initial implementation but on their ability to evolve, adapt, and continuously deliver value to stakeholders across the education ecosystem. Sustainability requires building systems that are technically robust, financially viable, and deeply embedded in the decision-making processes of education leaders at all levels. This final component focuses on creating the conditions for sustained growth, continuous improvement, and expanding impact over time. Successful sustainability strategies recognize that data systems are not static tools but dynamic platforms that must evolve with changing policy priorities, technological capabilities, and stakeholder needs. States that achieve long-term success invest in building strong user bases, establishing reliable feedback mechanisms, and creating organizational capacity for ongoing innovation and improvement.

**Key Priorities:** fostering continuous improvement, building strong feedback loops, addressing system stress points, and equipping stakeholders with the resources and training they need to be successful.

## 10.1 ENGAGE STAKEHOLDERS ON WAYS TO LEVERAGE DATA AND GATHER FEEDBACK

The foundation of any sustainable data system is an engaged, knowledgeable user base that sees clear value in the system and actively advocates for its continued development and support. This community includes not just technical users who interact with the system on a regular basis, but also decision-makers who rely on the insights generated by the system.

The following are ways to actively engage your user base:

- **Create feedback loops** with users to identify challenges to using the system and emerging needs, both in terms of technical supports and improvements to data governance. Feedback from stakeholders might be collected in one or more of the following ways: formal surveys, usage analytics, help desk reports, or advisory committees.
- **Provide user training** for end-users to teach them how to interact with and leverage the data system for their specific needs. Examples of user training may include: developing user guides, case studies, webinars, pre-recorded demos, or offering regular office hours.
- **Streamline data collection and submission processes**—particularly for LEAs with limited capacity—make it easier for Data Providers to submit high-quality, accurate data often and on time. Lighten capacity by gathering feedback on the data system’s usability and work to implement changes to ease data submission protocols. Streamlining data collection, making it simple and efficient for data providers, will also improve data governance, yielding more consistent and higher quality data.
- **Provide dedicated technical support** to respond to and resolve technical difficulties for all stakeholders and end-users.

## 10.2 MAKING IMPROVEMENTS AND CONTINUING TO MODERNIZE THE SYSTEM

Over time, data systems will need to evolve to support emerging needs and take advantage of new state-of-the-art technical capabilities—enhancing functionality and potentially reducing costs. However, any changes must be carefully considered in order to balance stability for existing users. Successful states approach system evolution strategically, with clear priorities, processes, and communication for evaluating and implementing changes to the system.

### Scaling Infrastructure and Capabilities:

As data systems mature and user bases grow, technical infrastructure must scale to maintain performance and reliability. Regularly assessing system optimization and performance ensures that growth doesn't compromise user experience. The following are elements of your system that should be assessed regularly:

- Database optimization and the ability to handle increasing data volumes efficiently
- Query performance monitoring to identify and resolve bottlenecks in requesting data
- Load testing to ensure system stability during peak usage periods
- Capacity planning to proactively address infrastructure needs before performance degrades
- Data Governance to ensure data are accurate, securely managed, usable and trustworthy
- Security monitoring to detect whether systems are susceptible to common types of attacks, before they occur, and whether known system vulnerabilities are present
- Compliance monitoring to ensure ongoing adherence to privacy regulations and data protection standards

### Responding to Evolving Stakeholder Needs:

User needs evolve as their experience with data systems grows and as policy priorities shift. As users become more sophisticated, they often demand more powerful analytical tools, such as:

- Predictive modeling capabilities that help identify trends and potential issues before they become critical
- Increased functionality to explore, filter, and visualize data
- Export and integration options that allow users to incorporate system data into their own analytical workflows
- Mobile access for enhanced accessibility

### Incorporating New Data Sources and Standards:

The value of workforce data systems increases as additional data sources are incorporated—providing a more complete picture of educator supply and demand across the state. Successfully incorporating new data sources requires:

- User impact analysis to understand how new data sources will alter existing workflows and reporting
- Standards alignment to ensure new data adheres to proper formats, standards, and expectations for data quality
- Technical integration planning to determine how new data will be processed, stored, and linked to existing information
- Privacy and legal review to ensure new data collection complies with applicable regulations

**Key Lessons:** start with valuing the user, create a culture of continuous improvement, leverage a common data standard, invest in support infrastructure (e.g., trainings, documentation, user support channels, workshops), collect and incorporate stakeholder feedback, and demonstrate policy impacts and quality improvements to stakeholders.

## BRIGHT SPOT EXAMPLE



The [Texas Insight to Impact \(I2I\) Dashboard and Data Governance Committee](#) demonstrate how sustainable systems can expand their scope over time to provide greater value. Texas expanded their data-driven capabilities by leveraging a new technology, the Ed-Fi Educator Preparation Data Model, and pulling new data sources from EPPs to build their [Insight to Impact \(I2I\) Dashboard](#). The I2I Dashboard provides the Texas Education Agency (TEA)

and EPPs in Texas with valuable insights and data about their candidates and programs. The I2I Dashboard brings together 9 data sources with 300+ data elements (down to the candidate level) on weekly and annual data collection cycles—totaling roughly 60M+ data records. Through user-friendly, online dashboards, this data are available to all 120 EPPs, at no cost to them, offering a wealth of actionable data to help EPPs improve their processes and outcomes.

From the newly published Ed-Fi [SEA Playbook: Advancing Data Utilization to Support Educator Preparation and Teacher Workforce](#), “SEAs like the Texas Education Agency (TEA) have transitioned from a primarily compliance-focused approach to gathering EPP data to one that emphasizes continuous improvement. This shift reflects that focusing on educator needs better supports the needs of students. TEA has been at the forefront of this shift, integrating data-driven decision-making and support systems that encourage ongoing development and innovation. This evolution highlights the critical role SEAs play in actively supporting the growth and quality of educational systems.”

Rather than simply collecting data to meet regulatory requirements, Texas has created a system that actively supports program improvement and innovation. By helping EPPs understand and improve their outcomes, the I2I dashboard creates opportunities to enhance the state’s EPPs in evidence-based ways. Furthermore, Texas focused on creating immediate value for EPPs rather than simply meeting compliance requirements—this approach builds user advocacy supporting long-term sustainability.



**KEY  
RESOURCE**

## [Ed-Fi Educator Pipeline Use Cases & Tools](#)

# CLOSING

Robustly navigating and responding to the needs of our nation’s teacher workforce represents both a significant challenge and an unprecedented opportunity. Success requires viewing data systems not as solely technical projects with defined endpoints, but as evolving platforms for continuous learning and improvement. This perspective demands ongoing investment in stakeholder relationships, technical capabilities, and organizational learning. However, states that make these investments create systems that become increasingly valuable over time, generating insights and supporting decisions that improve outcomes for educators and students alike.

The examples and strategies outlined in this roadmap demonstrate the individual components that lead to enduring and thriving workforce data systems. States should choose to focus on the components that are most crucial at this point in time for them, treating this roadmap as a guide.

You do not have to navigate these decisions alone. The Beyond100K Solution Lab is a valuable resource, offering insights and lessons learned from those who have already faced similar challenges. Discovering unknowns on your own can be frustrating and costly, so finding trusted partners who understand the potential sticking points can be hugely beneficial. Conducting a cost-benefit analysis can help you weigh the trade-offs between contracting with a technical implementation partner vs. handling the work in-house. We recommend that any state—whether you’re just starting out or looking to improve an existing data system—seek out experienced partners who can guide you through each decision, considering the unique context of your state.

Finally, the ultimate measure of success will not be the sophistication of the technology or the completeness of the data, but the degree to which these systems help ensure that every student has access to effective, well-supported educators who can help them reach their full potential.

# ACKNOWLEDGEMENTS

## Report written by:

Ivy Sandberg, Chief Operating Officer at SETA-ED

## Report edited by:

Dr. Melina Uncapher, Founder and Chief Executive Officer of SETA-ED; Talia Milgrom-Elcott, Founder and Executive Director of Beyond100K; Maya Morales Garcia, Chief Program Officer at Beyond100K; and Meghan Grady Serrano, Manager of Strategic Initiatives at Beyond100K

## Report co-published by:

- Beyond100K
- New Mexico Public Education Department (NMPED)
- Arkansas Department of Education, Division of Elementary and Secondary Education

## A special thanks to the following contributors who provided guidance, feedback, and participated in interviews or the Beyond100K Solution Lab Implementation Boot Camps, including:

- Sharlee Crowson, Educator Workforce Data Manager, Arkansas Department of Education
- Molly Gold, MPP, Project Manager Education, National Conference of State Legislatures
- Dr. Lindsey Judd, Manager of Programs, Ed-Fi Alliance
- Dr. Kristin Kipp, Director of Research and Impact at the Colorado Department of Education
- Katy MacCornack, BS MA, Recruitment & Retention Specialist, New Mexico Public Education Department (NMPED)
- Dr. Patricia Saenz-Armstrong, Head of Data and Research, Craft Education
- California Department of Education
- Edifying Teachers
- Emerson Collective
- Hawaii Department of Education
- Learning Policy Institute
- Maryland Longitudinal Data System Center
- The Missouri Department of Elementary and Secondary Education
- MiSTEM
- National Council on Teacher Quality
- Southern Regional Education Board
- Texas Education Agency
- Vermont Agency of Education

# GLOSSARY OF TERMS

Term	Definition
API	<b>Application Programming Interface.</b> A set of protocols and tools that allows different software applications to communicate and exchange data.
Beyond100K	Beyond100K unites leading STEM organizations to co-develop and implement solutions that will end the STEM teacher shortage by 2043, especially for those most excluded from STEM opportunities. For over a decade, Beyond100K has served as a field catalyst for systemic progress in STEM education, mobilizing hundreds of organizations to achieve results no single entity could accomplish alone. Since 2023, our 200+ partners have prepared 33,000 STEM teachers and supported nearly 120,000 more, impacting as many as 11 million students nationwide. See: <a href="https://beyond100k.org">beyond100k.org</a> .
Bootcamp Series	A series of tailored support opportunities to customize and apply the Roadmap components, and hear success stories from implementers in the field.
Data Governance	A system of policies, processes, roles, and standards that an organization uses to manage its data assets, ensuring they are accurate, secure, accessible, and compliant with regulations
Data Providers	Stakeholders who contribute data, such as Educator Preparation Programs (EPPs), teacher license boards, and Local Education Agencies (LEAs)/school districts.
Data Standard	A set of standardized definitions, formats, and coding schemes that enable seamless data integration and analysis across diverse sources.

Term	Definition
Data Stewards	Stakeholders (e.g., SEA and technical staff, Managed Service Providers) who collect, join, and manage centralized data stores, and define the rules and operating procedures for the centralized data system, including data governance, privacy, and security.
Data Users	Stakeholders who leverage data to inform data-driven research and/or policy initiatives, including Data Providers, Data Stewards, external partners, researchers, and policymakers.
Ed-Fi Alliance	The Ed-Fi Alliance is a nonprofit organization dedicated to improving educational outcomes by enabling secure, seamless data interoperability across K–12 education systems. Founded in 2012 as an initiative of the Michael & Susan Dell Foundation, the Alliance develops and maintains the Ed-Fi Data Standard—a common language that allows disparate education technology systems to exchange information efficiently and securely. See: <a href="https://www.ed-fi.org/">https://www.ed-fi.org/</a>
EPP	<b>Educator Preparation Program.</b> Includes higher education institutions, residency programs, or apprenticeship programs that prepare teacher candidates. They are a type of Data Provider.
LEA	<b>Local Education Agency.</b> A school district or similar local entity that provides data, often via HR systems, employment records, and student information systems (SIS). A type of Data Provider.
MSP	<b>Managed Service Provider.</b> Specialized vendors or technical consultants that a state may partner with to manage the data system, especially for states with limited internal technical capacity.

Term	Definition
NIST	<p><b>National Institute of Standards and Technology.</b> A U.S. government agency that publishes standards and guidelines for information security and data management, which states may reference for technical security requirements (Component 5.2) and data governance frameworks.</p>
ODS	<p><b>Operational Data Store.</b> A popular relational database and companion API harmonized within the Ed-Fi Data standard, often used by states for centralized data storage.</p>
SEA	<p>State Education Agency. The state-level agency that stewards the data and is responsible for managing the data system. Data Stewards and the Technical Team are often housed here.</p>
SETA-ED	<p><b>Scientific and Engineering Technical Assistance for Education.</b> SETA-ED is a scientific and technical research partner that provides white-label support to education organizations through scientific and engineering technical assistance via an interdisciplinary team of educational neuroscientists, learning engineers, data scientists, and data privacy and ethics experts.</p>
SIS	<p><b>Student Information System.</b> A source system (used by LEAs/school districts) where teacher workforce data, such as employment or student records, originates.</p>
TPDM	<p><b>Teacher Preparation Data Model.</b> An extension to the Ed-Fi standard that numerous states use to link EPP (Educator Preparation Program) data in their teacher data workflow.</p>

Term	Definition
UAT	<p><b>User Acceptance Testing.</b> An engagement strategy where primary users are involved in testing new features or the overall system before rollout to ensure it meets their needs and is user-friendly.</p>
ODS	<p><b>Operational Data Store.</b> A popular relational database and companion API harmonized within the Ed-Fi Data standard, often used by states for centralized data storage.</p>
SEA	<p>State Education Agency. The state-level agency that stewards the data and is responsible for managing the data system. Data Stewards and the Technical Team are often housed here.</p>
SETA-ED	<p><b>Scientific and Engineering Technical Assistance for Education.</b> SETA-ED is a scientific and technical research partner that provides white-label support to education organizations through scientific and engineering technical assistance via an interdisciplinary team of educational neuroscientists, learning engineers, data scientists, and data privacy and ethics experts.</p>
SIS	<p><b>Student Information System.</b> A source system (used by LEAs/school districts) where teacher workforce data, such as employment or student records, originates.</p>
TPDM	<p><b>Teacher Preparation Data Model.</b> An extension to the Ed-Fi standard that numerous states use to link EPP (Educator Preparation Program) data in their teacher data workflow.</p>

# APPENDIX

## APPENDIX A.

### ROADMAP PRIORITIZATION STRATEGIES FOR STATE LEADERS

---

This guide supports you in how to prioritize where to start on the **Teacher Workforce Data System Roadmap** and then how to move through the work. Use it alongside the Companion Guide’s (Appendix B) self-assessments and planning tables.

#### 1) How to Move Through the Work (Decision Rule)

This step gives you a simple way to choose where to start: use the 2-out-of-3 rule—urgency, data readiness, and stakeholder will. Rate all the components in the Roadmap and select the component that scores highest on at least two factors. Finally, log your choice and why in the Companion Guide’s “Next Steps” table so the team leaves with clarity.

- **Urgency:** Which problem, if answered, changes decisions this term?
- **Data readiness:** Which question has usable data today (per your inventory)?
- **Stakeholder will:** Where do you already have the coalition and approvals?
- Pick the option that wins 2 of the 3. Document the rationale in your planning table.

#### 2) Prioritization Lanes (Impact × Effort)

Use this matrix to sort your potential actions from the Roadmap into **Fast Wins, Enablers, and Big Bets**. The goal is to start at least one Fast Win right away, while planning one Enabler and keeping one Big Bet in view—balancing momentum with foundation-building.

- **Fast Wins:** High impact, low effort (start immediately)
- **Enablers:** Medium/high impact, medium effort—often prerequisites
- **Big Bets:** Highest impact, higher effort/complexity—plan and fund

# APPENDIX

## APPENDIX A.

### ROADMAP PRIORITIZATION STRATEGIES FOR STATE LEADERS

---

This guide supports you in how to prioritize where to start on the **Teacher Workforce Data System Roadmap** and then how to move through the work. Use it alongside the Companion Guide’s (Appendix B) self-assessments and planning tables.

#### 1) How to Move Through the Work (Decision Rule)

This step gives you a simple way to choose where to start: use the 2-out-of-3 rule—urgency, data readiness, and stakeholder will. Rate all the components in the Roadmap and select the component that scores highest on at least two factors. Finally, log your choice and why in the Companion Guide’s “Next Steps” table so the team leaves with clarity.

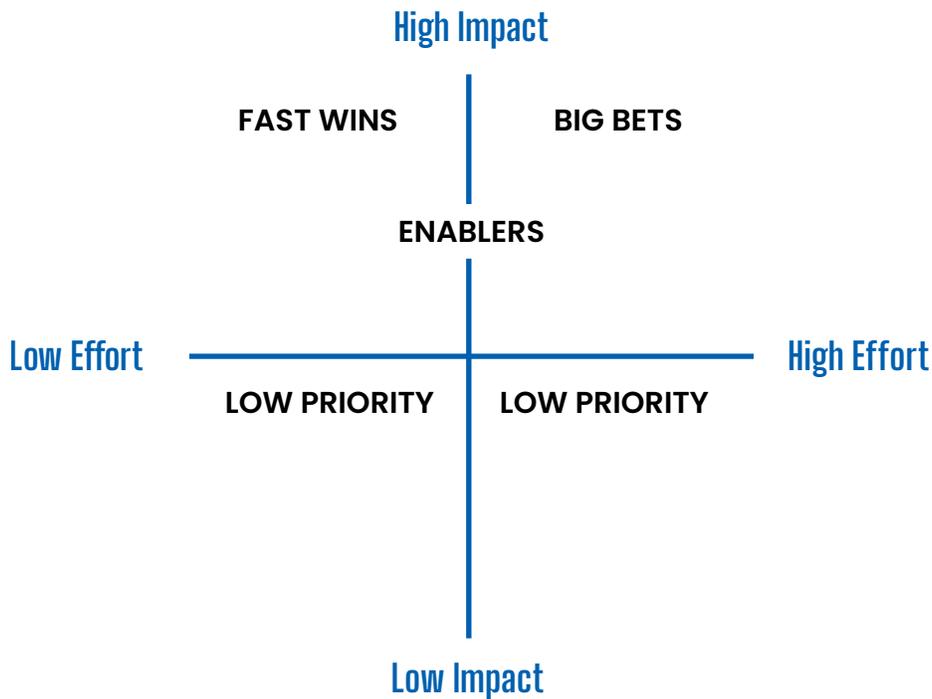
- **Urgency:** Which problem, if answered, changes decisions this term?
- **Data readiness:** Which question has usable data today (per your inventory)?
- **Stakeholder will:** Where do you already have the coalition and approvals?
- Pick the option that wins 2 of the 3. Document the rationale in your planning table.

#### 2) Prioritization Lanes (Impact × Effort)

Use this matrix to sort your potential actions from the Roadmap into **Fast Wins, Enablers, and Big Bets**. The goal is to start at least one Fast Win right away, while planning one Enabler and keeping one Big Bet in view—balancing momentum with foundation-building.

- **Fast Wins:** High impact, low effort (start immediately)
- **Enablers:** Medium/high impact, medium effort—often prerequisites
- **Big Bets:** Highest impact, higher effort/complexity—plan and fund

# IMPACT X EFFORT MATRIX



### 3) Dependency Map (Do this before that)

Now let's think about any blockers. Think of this table as your "gate check." If a prerequisite in the left column isn't in place, tackle it before—or in parallel with—anything it unlocks. This helps avoid rework and ensures each investment builds on a solid base.

Prerequisite	Then unlocks
Define 4–6 key terms (C4.2)	Consistent vacancy/retention metrics; apples-to-apples comparisons
Basic data inventory (C2.1)	Gap analysis, targeted collection (C2.2, C6.1)
RBAC + MFA (C5.2)	Secure sharing of dashboards and NLQ pilots (C8)
L1 validation rules (C7.1)	Reliable dashboards & automated reports (C8.1, C8.2)
Choose collection mechanism (C6.1)	Repeatable data flows and longitudinal linking (C7.2)

#### 4) Choose Your Starting Pathway

We identified three pathways—**Vacancy Response, Pipeline & EPP Outcomes, and Governance & Privacy First**— from patterns we saw across states in the field scan and are built from groupings of Roadmap components. The listed triggers are common conditions that signal when a pathway is the best starting point (e.g., widespread vacancies signal “Vacancy Response” as the right focus). Choose the pathway whose triggers most closely match your current situation, your top workforce questions, and the momentum you already have (per your Companion Guide self-assessment). Then, document the pathway’s first action, owner, and timeline in your Companion Guide.

A. **Vacancy Response Track** — define vacancy; collect monthly vacancy file; publish a filterable view.

- Triggers: widespread vacancies; leadership urgency; accessible HR extracts.

B. **Pipeline & EPP Outcomes Track** — compile EPP → license → placement → retention; publish outcomes brief.

- Triggers: strong EPP partnerships; policy focus on preparation quality; identifiers available.

C. **Governance & Privacy First Track** — stakeholder plan, DSAs/MOUs, RBAC tiers, privacy-by-design.

- Triggers: sensitive data concerns; fragmented governance; audit findings.

#### 5) 90-Day Plan (Three Sprints)

This step helps you deliver visible progress fast: each sprint should produce a tangible artifact (like a definitions one-pager, functioning RBAC, or a live dashboard) that you can show to leadership and stakeholders. Use the Companion Guide to capture actions, owners, and blockers for each sprint, and schedule brief check-ins at the end of each sprint to review progress, remove roadblocks, and adjust priorities. Keeping the scope small but visible ensures momentum, builds confidence, and secures buy-in for the next phase.

- **Sprint 1 (Weeks 1–3):** Self-assessment, top 2–3 questions, key terms, initial data inventory.
- **Sprint 2 (Weeks 4–7):** Pick collection mechanism; implement L1 validations; stand up RBAC; first automated report.
- **Sprint 3 (Weeks 8–12):** Publish a basic dashboard for one question; document definitions; schedule the next collection cycle.

## 6) Fast Wins Catalog

Pick one or two Fast Wins that can be delivered in weeks and will unblock other work—these are your momentum-builders. Aim for actions that give leaders new visibility (like a simple vacancy snapshot) or remove a pain point (like automating a recurring report). Use the Companion Guide’s planning table to assign an owner, timeline, and needed support, and share the completed artifact widely to demonstrate quick progress.

- Publish a one-pager with standardized definitions for vacancy, shortage, retention, uncertified.
- Share a living data inventory sheet naming stewards and refresh cadence.
- Automate one recurring static report (e.g., monthly vacancy snapshot).
- Turn on role-based access with least privilege for current users.
- Create a simple intake → answer loop for data questions.

## 7) Big Bets with High Payoff

Select at least one Big Bet that will create significant, sustained improvement in your teacher workforce data system—such as linking EPP to retention data or piloting a secure NLQ tool. Identify all prerequisites (from the Dependency Map) and begin with a clear discovery phase, budget outline, and governance plan. Keep leadership informed with periodic progress updates, even before implementation, to maintain support.

- Longitudinal linking across EPP → license → placement → retention to evaluate pipeline effectiveness.
- Modern architecture + security hardening (cloud/on-prem, encryption, logging, incident response).
- Self-service analytics and a scoped NLQ pilot on de-identified aggregates.

## Additional Considerations: Metrics That Matter

Decide on a small set of measures now (e.g., time-to-answer, % of data passing validation, dashboard adoption rates) and baseline them immediately. Update monthly or quarterly, using trends to show improvement and to secure continued investment. Tie these metrics to the Roadmap’s progress rubrics so leadership sees a clear link between actions taken and system maturity.

- Time-to-answer for the top 2–3 questions
- Reduction in ad-hoc analyst tickets
- % providers passing L1 checks
- On-time submissions
- Dashboard adoption
- Policy actions informed by data
- Vacancy/retention trend visibility.

### Additional Resources: Leadership Talk Track (Use as a script)

Adapt this script with your own top workforce questions, chosen Fast Win, and Big Bet to present a compelling, time-bound plan to decision-makers/funders/stakeholders. Aim to deliver it in a way that emphasizes quick wins, prudent foundations, and strategic long-term gains. Use it repeatedly—in meetings, briefings, and written updates—to keep all stakeholders aligned and reinforce the sequence of work.

“We’re focusing on the 2–3 workforce questions that change decisions this term. We’ll deliver one fast win in 30 days, lay the foundation with validation and access controls, and plan a high-payoff big bet. This sequence reduces burden on providers, accelerates insight for leaders, and keeps privacy and security at the core.”

*Aligned to the Roadmap Components and Companion Guide planning tables.*

## APPENDIX B.

### COMPANION GUIDE: TEACHER WORKFORCE DATA SYSTEM STATE-ASSESSMENT & STRATEGIC PLANNING

---

This Companion Guide supports state education agencies in operationalizing the **Teacher Workforce Data System Roadmap**. Each section aligns with a Roadmap component and includes a self-assessment checklist, progress rubric, reflection prompts, a next steps planning table, and optional tips. This guide is designed to be editable and used during the Solutions Lab Bootcamp and future strategic planning efforts. For a summary table that consolidates your responses in this document into a one-page overview, please see Appendix C.

# COMPONENT 1. MAKE THE CASE & BUILD THE FOUNDATION

## Self-Assessment Checklist

- Do we have leadership buy-in for strengthening teacher workforce data?
- Do we have clearly communicated goals tied to improving teacher pipeline and retention?
- Have we engaged cross-functional teams across HR, data, and policy?

## Progress Rubric

Tier	Description
Emerging	Workforce data efforts are fragmented with limited visibility.
Developing	Some alignment exists between teams and goals are partially articulated.
Established	Clear vision and cross-agency collaboration is driving workforce data reform.

## Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

## Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 2

## DEFINE THE PROBLEM(S)

### Self-Assessment Checklist

- Have we defined our workforce challenges with supporting data?
- Do we disaggregate data by subject, grade, geography, and demographics?
- Are problems linked to actionable policies or initiatives?

### Progress Rubric

Tier	Description
Emerging	Problems are described anecdotally without data support.
Developing	Some data-informed problem definitions exist.
Established	Problems are data-defined and aligned with strategy.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 3. BUILD THE TEAM

## Self-Assessment Checklist

- Do we have a cross-functional team that includes technical, analytical, and user voices?
- Have we identified internal capacity gaps and external partners?
- Is there clarity on roles, responsibilities, and coordination mechanisms?
- Do we have formal policies for data use, sharing, and stewardship?

## Progress Rubric

Tier	Description
Emerging	Team roles are unclear or focused narrowly.
Developing	Some roles are defined, with partial coordination.
Established	A cohesive team with defined roles and governance is in place.

## Reflection Prompts

Use the space below to reflect on what’s working, what’s missing, and what support might be needed.

## Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 4. STANDARDIZE DATA

## Self-Assessment Checklist

- Are key data elements standardized across systems?
- Do we have a working data dictionary?
- Have we aligned data definitions with federal and peer state guidance?

## Progress Rubric

Tier	Description
Emerging	Data elements are fragmented and inconsistent.
Developing	Some data definitions are standardized.
Established	Data infrastructure is harmonized and documented.

## Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

## Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 5.

## BUILD AND STRENGTHEN DATA INFRASTRUCTURE

### Self-Assessment Checklist

- Have critical architecture decisions been made around hosting, data storage, processing capabilities, and integration approaches?
- Are data privacy and security protocols clearly defined and documented?
- When possible, have data privacy and security protocols been automated?

### Progress Rubric

Tier	Description
Emerging	Critical architectural decisions are being considered.
Developing	Protocols for data storage, maintenance, security and privacy are in active development.
Established	Robust infrastructure and governance structures for ensuring privacy, security, accountability and trust are in place.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 6.

## DESIGN DATA COLLECTION MECHANISMS

### Self-Assessment Checklist

- Do we collect teacher workforce data systematically and regularly?
- Are collection tools user-friendly and accessible to data providers?
- Is there a feedback loop to improve collection processes?

### Progress Rubric

Tier	Description
Emerging	Data is collected inconsistently or burdens providers.
Developing	Standardized collection tools are being introduced.
Established	Efficient, equitable mechanisms support accurate data flow.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 7.

## DESIGN THE DATA PIPELINE

### Self-Assessment Checklist

- Do we have automated or semi-automated processes for ingesting, cleaning, and storing data?
- Are validation and quality checks part of the pipeline?
- Do we track data flow from source to reporting?

### Progress Rubric

Tier	Description
Emerging	Pipeline processes are manual and error-prone.
Developing	Some automation and data quality steps are in place.
Established	A well-functioning pipeline ensures integrity and traceability.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 8.

## BUILD SYSTEMS TO USE THE DATA

### Self-Assessment Checklist

- Do end users have access to dashboards, reports, or tools that meet their needs?
- Are insights regularly shared with decision-makers?
- Do outputs enable actionable responses to workforce challenges?

### Progress Rubric

Tier	Description
Emerging	Limited or static reporting is in use.
Developing	Basic tools exist but require refinement.
Established	Dynamic systems drive regular, data-informed decisions.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 9.

## CREATE THE CONDITIONS TO SUPPORT USE

### Self-Assessment Checklist

- Are users trained and supported to interpret and act on data?
- Is there a culture of inquiry and collaboration across teams?
- Do we regularly incorporate feedback from end users?

### Progress Rubric

Tier	Description
Emerging	Data use is limited to a few technical staff.
Developing	Some training and collaboration structures exist.
Established	Widespread, supported use of data is part of daily practice.

### Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

### Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# COMPONENT 10. ENSURE SUSTAINABILITY

## Self-Assessment Checklist

- Do we have a long-term vision and funding strategy for workforce data work?
- Are efforts embedded in broader state priorities?
- Are we monitoring and evolving systems over time?

## Progress Rubric

Tier	Description
Emerging	Initiatives are grant-dependent or short-term.
Developing	Some sustainability plans are in development.
Established	Workforce data efforts are institutionalized and future-ready.

## Reflection Prompts

Use the space below to reflect on what's working, what's missing, and what support might be needed.

## Next Steps Planning Table

Action Step	Owner	Timeline	Support Needed	Priority

# APPENDIX C.

## TEACHER WORKFORCE DATA SYSTEM STATE-ASSESSMENT OVERVIEW

This one-page summary is designed to help state data leaders quickly assess their current status across the 10 key components of the Teacher Workforce Data System Roadmap. Use this guide to identify which elements your state has in place, what’s in progress, and where there are opportunities for development. This tool serves as a launchpad for deeper planning during the Solutions Lab Bootcamp.

Roadmap Component	We Have This (✓)	We’re Working on This (Σ)	We Need This (X)	Notes / Next Steps
1. Make the Case & Build the Foundation				
2. Define the Problem(s)				
3. Build the Team				
4. Standardize Data				
5. Build and Strengthen Data Infrastructure				
6. Design Data Collection Mechanisms				
7. Design the Data Pipeline				
8. Build Systems to Use the Data				
9. Create the Conditions to Support Use				
10. Ensure Sustainability				