



NATIONAL STRATEGIC PLANNING
& ANALYSIS RESEARCH CENTER

STATE LONGITUDINAL DATA SYSTEM GOVERNANCE

This document contains confidential information constituting the protected intellectual property of the National Strategic Planning & Analysis Research Center (nSPARC) at Mississippi State University and may not be copied, distributed, or otherwise reproduced in any manner without the prior written consent of nSPARC at Mississippi State University.

State Longitudinal Data System Governance

Over the last few years, the federal government and states have made substantial efforts to develop and establish statewide longitudinal data systems (SLDS). The objective is not to create single data systems but to facilitate the use of data from multiple sources. To this end, the federal government alone has invested more than \$600 million to assist states with integrating early childhood, K-12, two-year college, university, and workforce data. States are in different stages of development and implementation of their system, but they all face the same question: how will the system be governed once it is completed? Addressing this question is critical because no single agency or entity has ownership of the system, but they all have an equal stake in its success. Complex technical and compliance issues require a mutual and collective understanding of policies and procedure as they relate to the very existence of these state systems. The challenge for states is to create a governance policy that allows an SLDS to exist while protecting the interests of all parties involved.

Governance can be best defined as the process or processes by which an organization operates to maintain its existence. From an organizational perspective, governance is a holistic approach that considers all the key elements that defines the organization. Good governance is one that looks at one key element in the context of the others as a whole. In the case of the SLDS, much attention has been given to the use of the data and the preservation of confidentiality and privacy. While this is a key component of an SLDS, it is only one of several. To comprehensively address the issue of security and the preservation of privacy it must be done in the context of all the elements that make up the SLDS. An SLDS can typically be defined by its scope, stakeholders, use or application, operational capacity, leadership and accountability, and sustainability. A good SLDS governance, therefore, is defined by six building blocks:



Scope

The scope defines the purpose of the SLDS and provides the general framework for supporting and institutionalizing its use. There is a growing consensus that in order to improve the education performance of our students, we need better, timely information for better decision making. This principle is at the heart of the SLDS.

The growing knowledge-based economy has raised concerns about our ability to create a competitive workforce that is highly educated and technically skilled. As such, schools, colleges, and universities are expected to perform better than ever in helping students achieve education and skill standards aligned with workforce needs. Current statistics, however, reveal that our current educational system faces many challenges from early childhood through higher education. For example, many children are not kindergarten-ready, others fail to pass third grade, and an appreciative number never complete high school. Of those who graduate from high school, many are not college- or workforce-ready. More students than ever go to college, but at the same time, an appreciative number fail to graduate, and college graduates struggle to find jobs, as many are unprepared to meet educational and skill requirements.

The SLDS can help answer questions at the core of educational performance and effectiveness or outcomes (see Appendix A) and can fulfill the need to go beyond “simply collecting data.” There is a greater demand for analyzing longitudinal data over a longer time span, from early childhood to postsecondary education and the workforce, in order to answer questions about education quality and outcomes. The risk in this type of effort is mindless empiricism – using data with no clear purpose or direction. Many believe that data hold the answers to all questions. This is a false assumption because data only have meaning when they measure what we intend to measure.

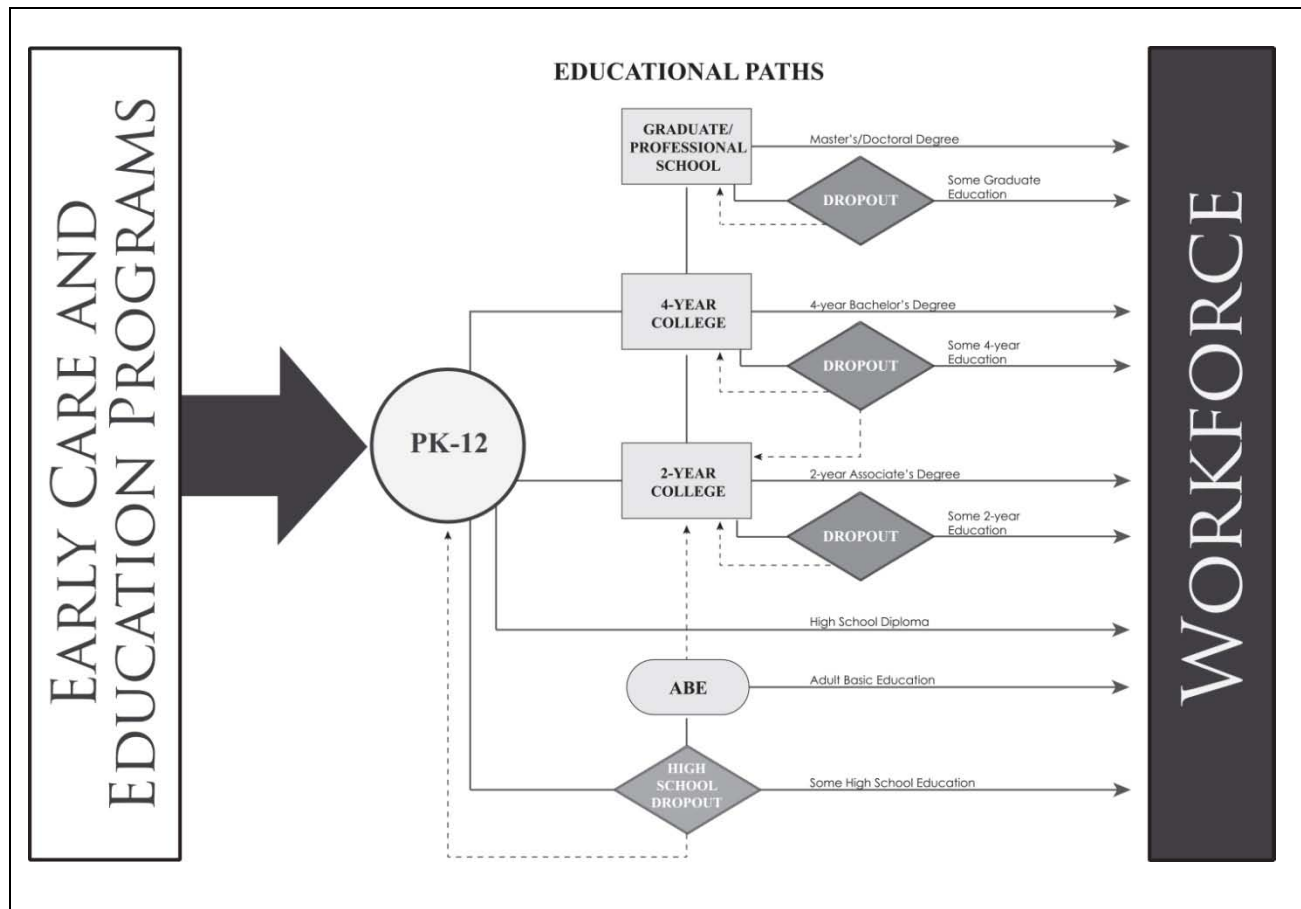
The focus here is the quality of educational outcomes, so the first step is to clearly define specific outcomes along educational pathways. The first pathway is to graduate from high school. Upon achieving the outcome of high school graduation, individuals can follow two pathways. They can gain full-time employment, or they can continue into postsecondary and receive some type of college credential. In the latter pathway, the next step is either earning a graduate or professional degree or gaining employment that requires some type of college credential. While these are the expected outcomes that define typical educational career pathways, many students fail to achieve the intended outcome. That is, many do not graduate from high school, many attending college do not earn a degree, and many are unprepared for the workforce and end up unemployed, underemployed, or out of the workforce altogether.

Outcomes outside conventional education pathways include completion of workforce training programs. As part of lifelong learning, many dislocated workers and incumbent workforce engage in education and skill development throughout their careers. These outcomes include completion of training and returning to school. These pathways are represented by the conceptual framework in Figure 1 (on next page).

This framework is at the heart of the elements required by the America COMPETES Act (see Appendix B) and guides the development of a detailed data model for the SLDS project. This

model helps to organize data in a way that allows us to, at all times, determine who succeeds and who fails and why. It also helps assess the quality of educational outcomes, as we will be able to determine how knowledge, content, and skills, along with teacher qualifications and teaching strategies, helps achieve intended outcomes along educational and career pathways.

Fig. 1: Outcomes along Educational Pathways



Data Stakeholders

A data stakeholder is an individual or organization that could affect or be affected by information generated from the SLDS. Obvious stakeholders include those who directly contribute data to the SLDS, such as state agencies, institutions of higher learning, and local school districts. Within these organizations, there are other stakeholders such as program subject matter experts, data subject matter experts, superintendents, principals, teachers, and workforce training providers. Other stakeholders are less obvious but can be clearly impacted by information generated from the SLDS. These stakeholders might include the general public, media, local leaders, policymakers, legislators, the Governor’s Office, and, of course, parents.

Applications

To avoid any confusion or misuse, the SLDS should only be used for activities directly aligned with its scope. The SLDS is not an accountability tool; it is a performance-based management tool for better decision making. The primary application of the SLDS is to generate timely, accurate, and policy-compliant information from data collected and stored by different local and state educational and workforce entities for the purpose of conducting activities within the scope, such as strategic planning, resource allocation and alignment, program evaluation, and scientific research.

Operational Capacity

Operational capacity is related to components needed to fulfill the scope of the SLDS while ensuring individual privacy and confidentiality of data. Operational capacity is established by data and technical expertise, legal and compliance expertise, formal agreements, policies and procedures for the data life cycle, research expertise, and the creation of a Center of Excellence to carry out all activities that define the operational capacity of the SLDS. A good guideline to assess SLDS operational capacity is the checklist from the Privacy Technical Assistance Center (PTAC) of the U.S. Department of Education (see Appendix C).

Data and technical expertise. The development and operation of an SLDS requires staff with significant expertise and experience in data analysis and database administration. This expertise is critical because data come in many forms and formats. Technical expertise must also include staff with information technology (IT) experience to manage hardware and secure networks. Programming and software development expertise is required to build the capacity for the development and establishment of the SLDS, as many off-the-shelf solutions are not ideal, and states must build their own custom solutions.

Legal and compliance expertise. The development and operation of an SLDS requires staff with experience in legal and compliance issues around the use of sensitive data. FERPA does not prohibit the use of educational data as part of an SLDS. It simply requires that identity of students remain, at all times, protected. Expertise in this area can avoid unnecessary confusion and facilitate the generation of information consistent with the scope of the SLDS.

Formal agreements. Establishing memorandums of understanding (MOUs) is a large part of the operational capacity. An MOU must clearly specify the scope, the data to be shared, and roles and responsibilities in handling the data. MOUs clearly outline data ownership and stewardship. The MOU is a legal document that clearly states that the data contributors maintain ownership of their own data and that any third party can only act on their behalf.

Policies and procedures for the data life cycle. This process includes the development of policies and procedures for a wide range of activities that cover transferring, managing, archiving, and monitoring data and other actions related to the process of operating an SLDS.

The process begins with the secure *transfer of data* from data stakeholders to the state data clearinghouse as governed by MOUs. The data transfer process is managed through a highly

secure method of exchanging data between partners that meets industry standard compliance requirements. Data transfer is provided via a platform independent secure web-based file transfer gateway that contains technical safeguards for protecting data and controlling access to those data. Email communication is sent to document each transfer so that a record of the transfer activities can be maintained by the state data clearinghouse for auditing and monitoring purposes.

After the transfer process is completed, the transferred data are saved to secure servers and inventoried. The *data inventory process* is a business-oriented process that involves managing and tracking the flow of data assets transferred to the state data clearinghouse. Building a comprehensive data inventory requires the expertise of data management coordinators that understand the flow of data that come into the state data clearinghouse. The data inventory process itself includes identifying all the data sets that comprise the SLDS and summarizing the information contained within the various data sets. The inventory tracks where data are located in the clearinghouse, when data are received, and to whom the data belong. The data dictionary and program data mapping documents are two instruments that are important to the data inventory process. The data dictionary is a centralized repository of information about the data. The data dictionary includes metadata such as meaning, relationship to other data, origin, usage, and format. The program data mapping document provides information on how data fields relate to other data fields across multiple data sets. The program data mapping document is used to compare and contrast data elements collected across multiple agencies. The data inventory is one of the foundational elements of the overall storage security practice and ties into access policies.

Once data are inventoried, the *data cleansing* process begins. This phase is an integral part of data processing and maintenance and entails a process designed to detect and correct inaccurate or incomplete records to create consistent and reliable data that will be consistent with other data sets in the system without changing the meaning of the data. This process results in clean data that can be de-identified and migrated into the SLDS. The data cleansing process includes running data diagnostics that help ensure high-quality, accurate, and complete data are available from the data contributors. The cleansing process also includes addressing any errors or omissions in the data (missing data) and the validation of data coding and values for all included data fields. When the data cleansing process is complete, the data will be consistent with other data sets and ready for de-identification and anonymization.

The *de-identification and anonymization* of the data involves the removal or obscuration of information in individual records so that the remaining information does not identify any individual, and so that there is no reasonable basis to believe the information could be used to identify an individual. To help alleviate privacy concerns, data classified as containing personally identifiable information (PII) in the data inventory phase is de-identified or anonymized before it is included for use in the SLDS by reducing the number of variables on which a match to an individual might be made. All de-identification procedures are implemented with the appropriate physical, technical, and administrative protections in place to minimize risk and ensure regulatory compliance. Additional safeguards include only releasing information generated from the SLDS in tabular form and applying appropriate suppression

techniques when cell sizes fail to meet minimum reporting requirements. De-identification of the data is documented to the data stakeholder contributing the original data.

Access to data and information in the SLDS will be through the one-stop portal, which will be enhanced with a custom business intelligence (BI) system that taps into the physical model populated with de-identified data. Access will be role-based, and de-identified individual data points will never be accessible through the state data clearinghouse. Individual level records will be made available only to individuals, agencies, or organizations that can have access to these records. Approval must come from the data contributor as specified by the MOU. Everyone will have access to the system but with role-based access. Roles will be determined by the data stakeholders. The system will produce aggregate statistics but will not allow access to individual-level information or records.

The *auditing and monitoring* of data is essential to alleviating concerns about handling sensitive data and is critical to effective policy and practice in handling SLDS data assets. Auditing and monitoring data exists in many forms and includes data loss prevention technology, access control technology, and storage area security solutions. Strict controls on access and permission for data folders using security access control settings are employed along with state-of-the-art software designed to monitor access to sensitive drives, folders, and datasets. This helps to ensure that all actions are reviewed and any suspicious activity can be detected. These solutions are combined with integrated firewall and intrusion protection hardware and software to ensure both internal and external integrity.

Risk management and training are two integral components of a comprehensive data management and security program. These components require an appreciation of the risks associated with handling sensitive information and are tied directly to the other aspects of the data life cycle. The goal of an effective risk management and training program is to balance potential risks with the mission and scope of the SLDS. Thus, it requires protecting IT and data assets while allowing the SLDS to perform its primary function. As such, this is both a management and technical function of the SLDS. This function starts with training so that data stakeholders, data clearinghouse employees, and end users understand compliance requirements and embrace risk management responsibility. Transparency regarding potential and significant risks is essential to building a culture that inherently minimizes risks to handling sensitive data. To guard against unforeseen system damage or inability to access local systems, physical backups of all data in multiple secure locations must be maintained. Ideally, a "hot" disaster recovery site in another city, where all data and software are continuously replicated, should be established. In the event of a prolonged power outage or disaster, the SDLS would be able to seamlessly shift to another operating location. A recovery site combined with effective access and auditing protocols will significantly reduce the risks associated with operating an SLDS.

Research expertise. This component will require staff who can apply scientific techniques for producing technical reports and provide overall descriptive pictures of basic statistics and trends. It also requires expertise for conducting applied research that implements experimental design techniques to evaluate programs. Last but not least, scientists or faculty who engage in basic research to increase the stock of common knowledge may produce scholarly works published in scientific journals using the SLDS.

SLDS Center of Excellence. SLDS initiatives often involve the creation of a Center of Excellence to support the development and establishment of the SLDS. The Center of Excellence acts as the state data clearinghouse. The state data clearinghouse is an organization designed to increase flexibility, streamline resources, share best practices, and establish common processes and standards for data security and integration. The clearinghouse becomes a place for sharing resources with all data stakeholders; defines the technical standards and process for data integration and data governance; and provides technical and scientific resources aligned with the scope for the appropriate use of the SLDS.

As a general approach, a successful data clearinghouse model must include a human dimension in the data life cycle, from transferring data to uploading data into the SLDS. The human dimension emphasizes the appropriate use of the data relative to the scope and data stakeholders rather than forcing data within preexisting applications.

Public universities provide an ideal setting for the establishment of a state data clearinghouse. Public universities, as state educational entities, provide a variety of expertise in conducting scientific research engaging with data stakeholders and fit easily within the scope of the SLDS. Universities have policies, procedures, and resources in place to handle sensitive data and protect human subjects. The culture of protecting individual identity is entrenched in the work conducted at public universities.

Leadership and Accountability

At the end of the day, governance is about leadership and accountability. Any initiative that does not clearly specify who assumes leadership roles has no clear accountability. The SLDS must have leadership that reflects the interest of the whole state rather narrowly focused or special interests. Creating statewide leadership provides one point for accountability.

SLDS leadership assures that the system stays true to its intended scope. Strong statewide leadership also assures there is clear ownership of the SLDS as a state tool for improving the quality of educational outcomes. SLDS leadership does not assume data ownership, which remains with the data contributor, but it assumes ownership of the process to generate the information needed from multiple data sources. SLDS leadership does not manage stakeholders but establishes and maintains partnerships and manages communications with all parties about the importance of supporting and benefitting from the SLDS. A true SLDS leadership is one that provides guidance in identifying roles and responsibilities for each of the building blocks that define SLDS governance.

Sustainability

Sustainability is the viability of the SLDS over time. Sustainability is established by legal authority through executive orders or legislation, continuous system innovation, ongoing training, and availability of resources. In order to be sustainable, the system must be able to evolve in a way that meets the changing needs of the state. Similarly, a system can only be sustainable if it is used appropriately, and as such, training becomes a critical component of a

sustainability plan. Finally, resources to keep the system operational are necessary for long-term viability.

SLDS Governance Structure

The SLDS governance building blocks provide the foundation to implement solid SLDS governance. The structure that emerges from these building blocks is built around a framework based on four major elements (see Figure 2).

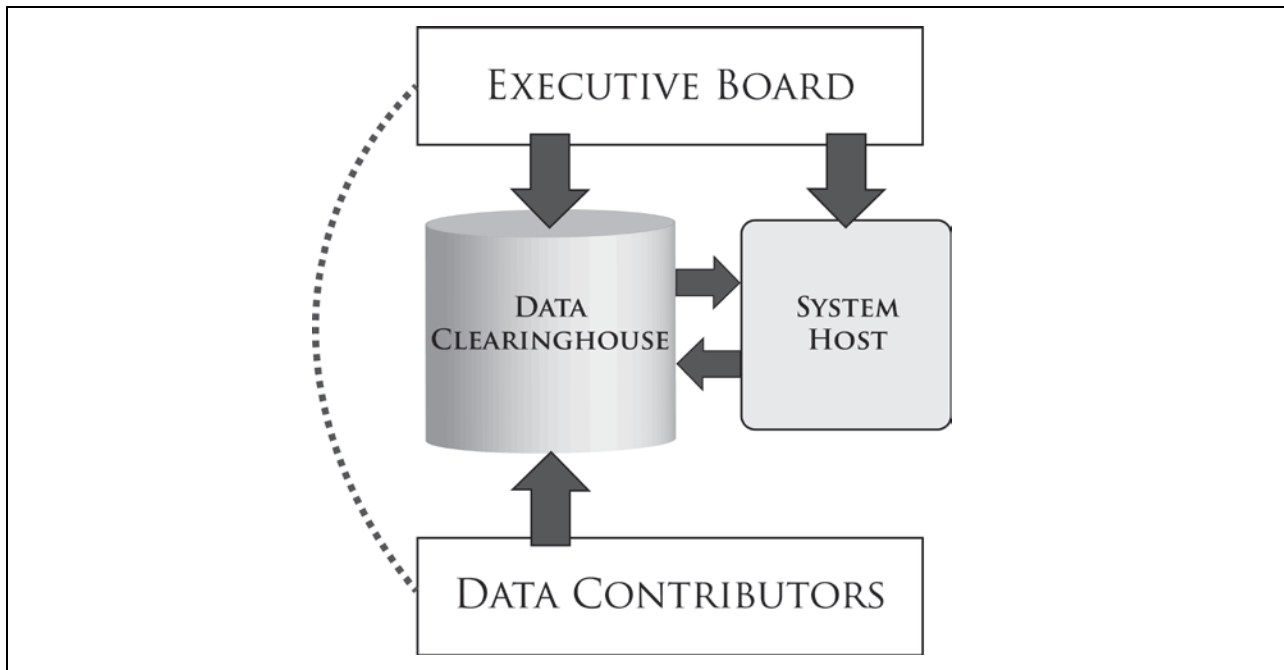
Executive board. This board embodies the state ownership of the SLDS and is responsible for safeguarding the integrity of the SLDS scope. The board will also provide leadership to keep data contributors engaged with the system and oversee the data clearinghouse and the system host.

Data contributors. Data contributors have responsibility of transmitting data to the data clearinghouse following the guidelines set forth in their MOUs. They also provide program subject matter experts and contribute to identifying critical data and information needs.

Data clearinghouse. The stewardship responsibility of the SLDS lies with this entity, which has the responsibility of establishing MOUs with the data contributors. The clearinghouse is also responsible for carrying out all activities associated with the data life cycle and providing support to maintain the SLDS.

System host. The system host is responsible for providing access to the SLDS and, therefore, hosting the hardware components of the system. It also provides the network and other infrastructure needed to access the one-stop portal through the custom BI.

Fig. 2: Governance Structure



APPENDIX A



www.erd.c.wa.gov
August 2, 2010

P-20 Research and Policy Questions

ESHB 2261 directs the Education Research and Data Center (ERDC) to identify the critical research and policy questions that ERDC intends to address. The list that follows is a collection of questions that will be used to steer data governance conversations and to inform the design of the Evergreen State P-20 longitudinal data system (ESP-20).

The questions have been compiled from a variety of sources and span the education sectors ranging from early learning, through K-12, post-secondary and into the workforce. The K-12 data governance committee, also established in ESHB 2261, was tasked with identifying critical research and policy questions but more specifically those that need to be addressed by the K-12 education data improvement system. A consulting group performed this work by reviewing national literature on longitudinal data systems and by interviewing and surveying representatives of stakeholder groups. Many of those questions identified as high priority for the K-12 system have been included in this ERDC collection of research and policy questions. Additional questions have been incorporated to cover the P-20 spectrum.

Themes of some questions appear across all education sectors: How does student achievement vary by characteristic, such as race/ethnicity, gender, income or financial aid status, participation in a particular program, or school attended. What factors are associated with differences in educational outcomes? Outcomes or student achievement may be measured by tracking whether a student progresses to the next education level, assessment scores, course completion, or employment.

Members of ERDC staff have considered numerous criteria in developing and prioritizing this list:

- *Expressed interest. Have members of the legislature, policy-makers, program staff, or the public expressed interest in this issue? Has the topic come up in policy or evaluation discussions?*
- *ERDC appropriate. Can ERDC add value by contributing information or analyses in this area? Does the question span educational sectors or does it involve longitudinal analyses? Agencies are generally the best source for sector-specific questions, although ERDC staff may contribute analytical resources to certain data topics.*
- *Size of student population being studied*
- *Overall system cost of the program or the students being studied*
- *Cost per student*
- *Data availability*
- *Ease of analysis*

In general ERDC will avoid narrow or local topics, but will address questions on a statewide level or will compare districts or geographies across the state. The list of possible questions is not intended to be exhaustive nor is it final. ERDC staff expect that the questions will evolve and priorities will change as issues arise and more data become available for analysis.

Question categories:

- Bold: an ERDC priority
- Question ERDC intends to address, but lower priority
- Data are not yet available (Bold indicates will be a priority when data become available)
- Question that is more appropriately answered "within sector", though ERDC could conduct analyses or contribute.

Student Profile

Who are the students? What are their characteristics?

- What are the demographic, mobility, program, class, grade, and course-taking profiles of students who do and do not achieve and what are their outcomes?
- Are students working while in school? What are the characteristics of working students?
- How many students receive financial aid and what are their characteristics?
- What are the characteristics and academic profile of students who are new to the state and to specific districts?
- To what extent are Washington bachelor's degree holders going on to graduate/professional school?
- What are the characteristics of graduate or professional school students? (baccalaureate majors, race/ethnicity, gender, occupation, etc)
- What are the characteristics of students in a school who have been involved in discipline incidents, suspended, expelled, or dropped out of school?
- What percent of children entering kindergarten are kindergarten-ready? How does kindergarten readiness relate to a child's performance in subsequent grades?
- Compared to state trends, what are the variations in district/school enrollment trends at different grade levels by gender, ethnicity, eligibility for free/reduced lunch, students in special education, students in ELL programs, and combinations?
- How have district/school subgroup attendance patterns changed at different grade levels?
- What are the characteristics of high attendance and low attendance students by school, grade level, and subgroup?
- What is the previous academic and attendance record of students who are new to the district, by school?
- What percentage of students transfer in or out at specific times of the school year by subgroup and where do they go?

Quality / Achievement
What are students doing? How well?
Course-taking, assessments, tests, etc

- How do the performance profiles of high mobility students compare to those of other students, e.g., attendance, proficiency, graduation, post-secondary enrollment?
- Is there a relation between college major and time-to-degree?
- How do need-based aid recipients compare to non-need-based aid recipients with respect to academic preparation in high school, remedial needs at postsecondary institutions, degree completion, time to degree, etc.
- What are the early indicators of success or failure for students in an elementary school, i.e., what is the K–3 profile of students who either succeeded or failed?
- What are the attendance patterns and proficiency levels of students who drop out by subgroup?
- How have individual district/school subgroup participation rates in AP, IB, SAT, and ACT exams changed and how do they compare to similar districts/schools?
- What is the relation between absence and performance on state assessments for different subgroups?
- How have individual district/school subgroup participation rates in advanced middle school courses changed and how do they compare to similar districts/schools?
- How do district/school changes in the percent of students who pass AP courses and ACT, SAT, and IB exams compare to state trends?
- What is the grade to grade progress of student subgroups on the state assessments in reading and mathematics, i.e., what percent of students initially below proficient reach proficiency and what percent either maintain or lose proficiency over time?
- Are English Language Learners making progress to attain English proficiency, by school and district?
- What is the relation between grades and performance on state assessments?

Transition / Advancement
Outcomes: Do students continue on education path?
Graduation rates, dropouts, retention, employment

- How are students from specific high schools performing at the post secondary level, and what are the strongest predictors of post secondary success, i.e., what are the high school profiles of students who succeed at the post secondary level?
- What are the education and workforce outcomes of low-income students (Free or Reduced Price Lunch-eligible students)? What are their postsecondary financial aid profiles?

- What are the outcomes and characteristics for those who drop out of high school before earning a diploma? How many re-enter? Get a GED? Enroll in the community and technical college system? Enter the workforce?
- What are retention rates of baccalaureate students, by institution and student characteristics?
- To what extent does swirling (transfer between institutions) occur among Washington baccalaureate and community and technical colleges?
- Where are baccalaureate graduates after graduation? What are their employment characteristics? Are they enrolled in graduate / professional school?
- What are the differences in time-to-degree for transfer students compared to students entering from high school?
- Do students in dual credit programs (Running Start, College in the High School, AP, IB) complete their baccalaureate degrees faster than students who don't enroll in these acceleration programs?
- On average, how much longer do part-time college students take to complete their degree than full-time students?
- Compared to full-time college students, how many more part-time college students drop out before graduation?
- What factors or characteristics are associated with reverse transfer college students (4-year to 2-year)? With horizontal transfer college students (2-year to 2-year; 4-year to 4-year)?
- What are the education and workforce outcomes of foster youth?
- What are the K-12 outcomes for ECEAP and Head Start participants? To what extent are they different?
- Do the effects of early childhood interventions "fade out" later?
- What are the strongest elementary school indicators of success or failure in the transition from elementary school to middle school, i.e., what are the elementary school profiles of students who succeed or fail in middle school?
- What are the strongest middle school indicators of success or failure in the transition from middle school to high school, i.e., what are the middle school profiles of students who either succeeded or failed?
- What are characteristics and outcomes of transfer students (and reverse transfer students)?
- What is the distribution of dropouts over the school year by subgroup and which groups have the highest dropout rates?
- How do increases or decreases in district/school dropout rates by subgroup compare to state dropout rates and dropout rates in similar districts/schools?
- How do district/school NCLB graduation rates for subgroups compare to state graduation rates and graduation rates in similar districts/schools?

Effectiveness and Costs

Evaluation and comparisons of programs, schools, districts

- What are the characteristics of districts/schools that meet or do not meet accountability requirements, i.e., funding, programs and course offerings, average class size, staff allocations, and teacher qualifications?
- What are the characteristics of districts/schools that show the greatest success in helping low achieving students reach proficiency?
- What is the cost effectiveness of specific district/school programs, i.e., what are the per-pupil costs (personnel and program material costs) of programs that have improved the performance of specific subgroups?
- At the aggregate level, what are the expenditures (personnel and non-personnel) for the major expense categories defined by the district, i.e., regular education, special education, vocational education, administration, transportation, maintenance, etc.?
- What are characteristics of schools with high achievement and low achievement gaps for students?
- Which colleges do the best job in helping diverse subgroups attain a degree?
- What are the returns to education by level, degree program, curricular structure?
- What programs, services, and instructional models have shown the most success in improving the performance of students in special education and ELL programs in similar districts/schools?
- What reading and mathematics programs/interventions have shown the most success in increasing student proficiency at the elementary, middle, and high school levels in similar districts/schools?
- What dropout prevention programs have shown the most success in decreasing dropout rates in similar districts/schools?
- What are the characteristics of districts/schools that show the greatest success in improving the performance of students in special education and ELL programs?
- What is the instructional cost breakout by federal, state, and local revenues at the district, school, program, and classroom levels?

Teachers

Supply, distribution, retention, training

- What are the most common characteristics of the teacher workforce in schools that show the greatest success with students?
- What are the common characteristics of teachers who leave the teaching workforce? What are their subsequent employment characteristics?
- What are the differences in qualifications and experience of teachers across classrooms, i.e., is the quality of the teachers equitable across classrooms and different achievement levels?

- What are the characteristics of teachers who show the greatest success and least success in improving student achievement and outcomes?
- What are the characteristics of elementary classrooms (e.g., class size, student demographics, paraprofessional support) that show the greatest success in improving student proficiency?
- What were the pre-service programs of teachers who have high student success rates over time?
- Are National Board Certified Teachers more effective in improving student achievement in challenging schools than non-National Board Certified Teachers?

APPENDIX B



The Alignment of the DQC’s 10 Essential Elements with the America COMPETES 12 Elements

In 2007, through the America COMPETES Act (Public Law 110–69), the federal government codified the 12 elements of a P-16 education data system. Through the American Recovery and Reinvestment Act, the federal government reaffirmed its commitment to the development and use of these data systems by requiring states, as a condition of receiving State Fiscal Stabilization Funds, to commit to building a data system which consists of all the elements indicated in America COMPETES. The table below shows how the 12 elements that were detailed in America COMPETES are aligned with the DQC’s 10 Essential Elements.

DQC 10 Essential Elements	Comparable COMPETES Element
1. A unique statewide student identifier that connects student data across key databases across years	1. A unique statewide student identifier that does not permit a student to be individually identified by users of the system
2. Student-level enrollment, demographic and program participation information	2. Student-level enrollment, demographic, and program participation information
3. The ability to match individual students' test records from year to year to measure academic growth	6. Yearly test records of individual students with respect to assessments under section 1111(b) of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 6311(b))
4. Information on untested students and the reasons they were not tested	7. Information on students not tested by grade and subject
5. A teacher identifier system with the ability to match teachers to students	8. A teacher identifier system with the ability to match teachers to students
6. Student-level transcript information, including information on courses completed and grades earned	9. Student-level transcript information, including information on courses completed and grades earned
7. Student-level college readiness test scores	10. Student-level college readiness test scores
8. Student-level graduation and dropout data	3. Student-level information about the points at which students exit, transfer in, transfer out, drop out, or complete P-16 education programs
9. The ability to match student records between the P-12 and higher education systems	4. The capacity to communicate with higher education data systems
	11. Information regarding the extent to which students transition successfully from secondary school to postsecondary education, including whether students enroll in remedial coursework
	12. Other information determined necessary to address alignment and adequate preparation for success in postsecondary education
10. A state data audit system assessing data quality, validity and reliability	5. A State data audit system assessing data quality, validity, and reliability

APPENDIX C



PRIVACY TECHNICAL
ASSISTANCE CENTER

For more information please visit the Privacy Technical Assistance Center:
<http://nces.ed.gov/ptac>

Data Governance Checklist

Overview

The U.S. Department of Education established the Privacy Technical Assistance Center (PTAC) as a “one-stop” resource for education stakeholders to learn about privacy, confidentiality, and security practices related to student-level longitudinal data systems. PTAC provides timely information and updated guidance on privacy, confidentiality, and security practices through a variety of resources, including training materials and opportunities to receive direct assistance with privacy, confidentiality and security of information in longitudinal data systems. More PTAC information is available on <http://ed.gov/ptac>.

Purpose

The purpose of this checklist is to assist stakeholder organizations, such as state and local educational agencies, with establishing and maintaining a successful data governance program to help ensure the individual privacy and confidentiality of education records. Data governance can be defined as an organizational approach to data and information management that is formalized as a set of policies and procedures that encompass the full life cycle of data, from acquisition to use to disposal. This includes establishing decision-making authority, policies, procedures, and standards regarding data security and privacy protection, data inventories, content and records management, data quality control, data access, data security and risk management, data sharing and dissemination, as well as ongoing compliance monitoring of all the above-mentioned activities. Specific best practice action items about the key data privacy and security components of a data governance program are summarized below. This document focuses on data governance of kindergarten through grade 12 (K-12) data systems. Data governance of the systems spanning postsecondary education, as well as those including pre-school education, may involve additional considerations outside the scope of this list.

Data Governance Checklist

Decision-making authority

Assigning appropriate levels of authority to data stewards and proactively defining the scope and limitations of that authority is a prerequisite to successful data management.

- Has an organizational structure with different levels of data governance (e.g., executive, judicial, legislative, administrative, etc.) been established, and roles and responsibilities at various levels specified (e.g., governance committee members, technology leaders, data stewards, etc.)?
- Have data stewards (e.g., program managers) responsible for coordinating data governance activities been identified and assigned to each specific domain of activity?

- Are data stewards' roles, responsibilities, and accountability for data decision making, management, and security been clearly defined and communicated (to data stewards themselves as well as other relevant stakeholders)?
- Do data stewards possess the authority to quickly and efficiently correct data problems while still ensuring that their access to personally identifiable information (PII) is minimized in order to protect privacy and confidentiality?

Standard policies and procedures

Adopting and enforcing clear policies and procedures in a written data stewardship plan is necessary to ensure that everyone in the organization understands the importance of data quality and security—and that staff are motivated and empowered to implement data governance.

- Have policy priorities affecting key data governance rules and requirements been identified, and has agreement (either a formal agreement or a verbal approval) on priorities been secured from key stakeholders?
- Have standard policies and procedures about all aspects of data governance and the data management lifecycle, including collection, maintenance, usage and dissemination, been clearly defined and documented?
- Are policies and procedures in place to ensure that all data are collected, managed, stored, transmitted, used, reported, and destroyed in a way that preserves privacy and ensures confidentiality and security (this includes, but is not limited to maintaining compliance with the Family Educational Rights and Privacy Act [FERPA])?
- Has an assessment been conducted to ensure the long-term sustainability of the proposed or established data governance policies and procedures, including adequate staffing, tools, technologies, and resources?
- Does an organization have a written plan outlining processes for monitoring compliance with its established policies and procedures?
- Have data governance policies and procedures been documented and communicated in an open and accessible way to all stakeholders, including staff, data providers, and the public (e.g., by posting them on the organization's website)?

Data inventories

Conducting an inventory of all data that require protection is a critical step for data security projects. Maintaining an up-to-date inventory of all sensitive records and data systems, including those used to store and process data, enables the organization to target its data security and management efforts. Classifying data by sensitivity helps the data management team recognize where to focus security efforts.

- Does the organization have a current inventory of all computer equipment, software, and data files?
- Does the organization have a detailed, up-to-date inventory of all data elements that should be classified as sensitive (i.e., data that carry the risk for harm from an unauthorized or inadvertent disclosure), PII, or both?
- Have data records been classified according to the level of risk for disclosure of PII?
- Does the organization have a written policy regarding data inventories that outlines what should be included in an inventory and how, when, how often, and by whom it should be updated?

Data content management

Closely managing data content, including identifying the purposes for which data are collected, is necessary to justify the collection of sensitive data, optimize data management processes, and ensure compliance with federal, state, and local regulations.

- Does the organization have a clearly documented set of policy, operational, and research needs that justify the collection of specific data elements (e.g., what PII needs to be collected to successfully monitor a student's participation in and progress through the education system)?
- Does the organization regularly review and revise its data content management policies to assure that only those data necessary for meeting the needs described above are collected and/or maintained?

Data records management

Specifying appropriate managerial and user activities related to handling data is necessary to provide data stewards and users with appropriate tools for complying with an organization's security policies.

- Have mechanisms been put in place to de-identify PII data whenever possible (e.g., by removing all direct and indirect identifiers from PII)?
- Has the organization established and communicated policies and procedures for handling records throughout all stages of the data lifecycle, including acquiring, maintaining, using, and archiving or destroying data?

Data quality

Ensuring that data are accurate, relevant, timely, and complete for the purposes they are intended to be used is a high priority issue for any organization. The key to maintaining high quality data is a proactive approach to data governance that requires establishing and regularly updating strategies for preventing, detecting, and correcting errors and misuses of data.

- Does the organization have policies and procedures in place to ensure that data are accurate, complete, timely, and relevant to stakeholder needs?
- Does the organization conduct regular data quality audits to ensure that its strategies for enforcing quality control are up-to-date and that any corrective measures undertaken in the past have been successful in improving data quality?

Data access

Defining and assigning differentiated levels of data access to individuals based on their roles and responsibilities in the organization is critical to preventing unauthorized access and minimizing the risk of data breaches.

- Are there policies and procedures in place to restrict and monitor staff data access, limiting what data can be accessed by whom, including assigning differentiated levels of access based on job descriptions and responsibilities? Are these policies and procedures consistent with applicable local, state, and federal privacy laws and regulations (including FERPA)?
- Have internal procedural controls been established to manage user data access, including security screenings, training, and confidentiality agreements required for staff with PII access privileges?
- Are there policies and procedures in place to restrict and monitor data access of authorized users (e.g., researchers) to ensure the conditions of their access to data in the system are consistent with those outlined in the data governance plan, including which data elements can be accessed, for what period of time, and under what conditions?

Data security and risk management

Ensuring the security of sensitive and personally identifiable data and mitigating the risks of unauthorized disclosure of these data is a top priority for an effective data governance plan.

- Has a comprehensive security framework been developed, including administrative, physical, and technical procedures for addressing data security issues (such as data access and sharing restrictions, strong password management, regular staff screening and training, etc.)?
- Has a risk assessment been undertaken, including an evaluation of risks and vulnerabilities related to both intentional misuse of data by malicious individuals (e.g., hackers) and inadvertent disclosure by authorized users?
- Is a plan in place to mitigate the risks associated with intentional and inadvertent data breaches?
- Does the organization regularly monitor or audit data security?
- Have policies and procedures been established to ensure the continuity of data services in an event of a data breach, loss, or other disaster (this includes a disaster recovery plan)?
- Are policies in place to guide decisions about data exchanges and reporting, including sharing data (either in the form of individual records containing PII or as de-identified aggregate reports) with educational institutions, researchers, policymakers, parents, and third-party contractors?
- When sharing data, are appropriate procedures, such as sharing agreements, put in place to ensure that any PII remains strictly confidential and protected from unauthorized disclosure? Note that data sharing agreements must be authorized in applicable local, state, and federal privacy laws and regulations, such as FERPA. These agreements can only take place if data sharing is permitted by law.
- Are appropriate procedures, such as rounding and cell suppression, being implemented to ensure that PII is not inadvertently disclosed in public aggregate reports and that organization's data reporting practices remain in compliance with applicable local, state, and federal privacy laws and regulations (e.g., FERPA)?
- Are stakeholders, including eligible students or students' parents, regularly notified about their rights under applicable federal and state laws governing data privacy?

Please note that all recommendations included in this issue brief are intended to complement, not supersede, an organization's local security regulations and policies.

Glossary

Direct identifiers include information that relates specifically to an individual such as the individual's residence, including for example, name, address, Social Security Number or other identifying number or code, telephone number, e-mail address, or biometric record.

Education Records. include those records that are directly related to a student and are maintained by an educational agency or institution or by a party acting for the agency or institution. For more information, see the Family Educational Rights and Privacy Act regulations, 34 CFR §99.3.

Indirect identifiers include information that can be combined with other information to identify specific individuals, including, for example, a combination of gender, birth date, geographic indicator and other descriptors. Other examples of indirect identifiers include place of birth, race, religion, weight, activities, employment information, medical information, education information, and financial information.

Personally identifiable information (PII) includes information that can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other information. See Family Educational Rights and Privacy Act regulations (34 CFR §99.3) for a complete definition of PII specific to education data and for examples of education data elements that can be considered PII.

Sensitive data are data that carry the risk for adverse effects from an unauthorized or inadvertent disclosure. This includes any negative or unwanted effects experienced by an individual whose personally identifiable information (PII) was the subject of a loss of confidentiality that may be socially, physically, or financially damaging, as well as any adverse effects experienced by the organization that maintains the PII. See NIST, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, 2010 Special Publication 800-122, for more information.

Additional Resources

Investigating links to more detailed discussions about data governance will help an organization ensure the privacy, confidentiality, and security of its data.

The Family Educational Rights and Privacy Act (FERPA) Legislation (20 U.S.C. § 1232g; 34 CFR Part 99), FERPA Regulations: <http://www2.ed.gov/policy/gen/guid/fpco/pdf/ferparegs.pdf>

National Association of State Chief Information Officers (NASCIO) Enterprise Architecture publications - Data Governance series: Part I, II, and III (2008 and 2009):
<http://www.nascio.org/publications/pubsSubject2.cfm?category=14>

Statewide Longitudinal Data Systems (SLDS) Technical Brief 2. Data Stewardship: Managing Personally Identifiable Information in Electronic Student Education Records (NCES 2011-602):
<http://nces.ed.gov/pubs2011/2011602.pdf>

Supporting Data Use While Protecting the Privacy, Security and Confidentiality of Student Information:

A Primer for State Policymakers (2011): <http://dataqualitycampaign.org/files/DQC-Privacy-primer-Apr27.pdf>

PTAC's Data Governance and Stewardship Paper: <http://www2.ed.gov/policy/gen/guid/ptac/pdf/issue-brief-data-governance-and-stewardship.pdf>

About the Author

Dr. Domenico "Mimmo" Parisi is Professor of Sociology and Director of the National Strategic Planning & Analysis Research Center (nSPARC) at Mississippi State University. The major thrust of his research is in the area of workforce, economic, business and community development. He has conducted groundbreaking work in place-based economic and workforce development and created a performance management system to assist the state of Mississippi with the management and allocation of public resources. Dr. Parisi is extensively published with his main work appearing in technical reports, book chapters, policy briefs, and journals such as *Demography*, *Rural Sociology*, *Social Science Quarterly*, *Community Development Society*, *Society and Natural Resources*, and *Journal of Poverty*.

About nSPARC

The National Strategic Planning & Analysis Research (nSPARC) is a unique interdisciplinary center dedicated to excellence in research and economic and policy development. nSPARC houses and manages an education and workforce data warehouse for Mississippi. The warehouse includes records that can be linked across programs dating back to 2000 and has been used to generate outcome measures for educational attainment, employment, employment retention, and wages. The warehouse has also been used to track progress of students throughout PK-20 in the state. nSPARC is also the architect of and home to the State Integrated Longitudinal Education and Workforce Performance Management System, a tool that identifies best practices and tracks education and workforce outcomes.

