Data Governance and Stewardship

About PTAC

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

PTAC welcomes input on this document and suggestions for future technical assistance resources relating to student privacy. Comments and suggestions can be sent to PrivacyTA@ed.gov.

Purpose

Increased demand for high-quality education data and the resulting growth in the amount of individual student data collected and stored electronically by educational agencies has necessitated greater scrutiny of data management and protection practices. State and local educational agencies have expressed concerns about how to ensure data availability and quality while carefully preserving individual privacy. The purpose of this brief is to provide guidance on how to successfully manage complex data systems by establishing a comprehensive data governance approach. Data governance principles discussed in this paper apply to a large number of audiences and can be used to improve data management of systems spanning preschool through postsecondary education and into the workforce. The main audience is expected to be data stewards of kindergarten through grade 12 (K-12) data systems.

Establishing a comprehensive data governance program will help to ensure confidentiality, integrity, and availability of the data by reducing data security risks due to unauthorized access or misuse of data. Specifying standards, policies, procedures, and responsibilities regarding data ownership and data-related activities will help organizations to minimize any detrimental outcomes in the event of a data breach. The brief begins with an overview of data governance, then discusses the necessary components of a robust data governance program and reviews steps for implementing it. For more details on best practices in data governance, see PTAC's Data Governance Checklist.

Data Governance as the Operating Approach to Data Management

This section defines data governance and explains the importance of the concept, then summarizes the benefits of implementing a data governance system.

What is data governance? 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. The key steps to creating a robust data governance program include

- specifying organizational decisionmaking authority;
- specifying data standards and policies and procedures for guiding various data management processes, including data security and privacy protection, data quality control, and data dissemination activities;
- implementing these policies and procedures; and
- monitoring for compliance with the established standards and policies and procedures.

A data governance program includes

- protection of sensitive data;
- vulnerability assessment and risk management;
- enforcement of legal, regulatory, contractual, and architectural compliance requirements;
- identification of stakeholders, such as parents and administrators, their roles and responsibilities; and
- access management.

Why is data governance important? Proactive data governance is necessary to ensure confidentiality, integrity, accessibility, availability, and quality of the data. Establishing data governance is a critical task for any organization. It requires defining the organizational vision, policies, and practices; gaining support of the stakeholders; implementing the program; and monitoring its success. By clearly outlining policies, standard procedures, responsibilities, and controls surrounding data activities, a data governance program helps to ensure that information is collected, maintained, used, and disseminated in a way that protects the individuals' rights to privacy, confidentiality, and security, while producing timely and accurate statistical data.

What can be accomplished by implementing data governance? Data governance is a key factor for successful management of complex data systems. It enables organizations to deal more effectively with challenges related to data sharing, changes over time, and satisfying the needs of different stakeholder groups within and outside of the agency. The benefits of a proactive approach to data governance include

- improved data accuracy, achieved by scheduling regular data quality audits and using consistent data standards in variable naming and metadata categorization;
- improved data usability, resulting from monitoring data content for consistency with the organizational vision and stakeholders' needs;
- improved data timeliness, accomplished by avoiding unnecessary duplication of data collection efforts and reducing the work necessary to reconcile errors or discrepancies when merging or sharing data; and
- increased data security, gained by designing a comprehensive security plan and applying appropriate levels of protection to the data based on their level of sensitivity.

Data Governance Program Components

This section lists the ten key components of a comprehensive data governance program and summarizes critical issues that should be considered when designing the program. The first six components encompass the rules of engagement, the next three components deal with organizational bodies and individuals, and the last component describes processes needed to govern the data.

Policies and standards: To help ensure that the data governance program addresses organizational needs, it is necessary to clearly specify the "rules of engagement," or policies and standards that guide data governance program implementation. These include vision and goals of the organization regarding data standards, data management processes, decisionmaking jurisdiction, responsibilities, enforcement, and controls. The data governance approach should be consistent with the organization's overall mission and stakeholders' expectations. Program goals should be clearly stated, and it should be made clear how these goals address data content needs, what outcomes will be considered a success, and how the progress will be measured. Further, an organization should evaluate the resources required for the long-term sustainability of the program to

ensure it can sustain necessary levels of data quality and security over the entire data lifecycle. The main "rules of engagement" are as follows:

1. Mission and vision
2. Goals, governance metrics, success measures, and funding strategies
3. Data rules and definitions
4. Decision rights
5. Responsibilities and enforcement and compliance mechanisms
6. Security controls for risk management

Organizational bodies and individuals: The next three data governance components address the question of who is responsible for ensuring that the data governance program is implemented efficiently and effectively. It is critical to identify all relevant stakeholders, including data owners and users, and secure their feedback regarding policy priorities to ensure stakeholder buy-in and continuing support for the program. Key organizational bodies, including a data governance committee, should be created and individual data stewards should be assigned by the committee, with their roles and responsibilities explicitly outlined in the written data governance plan (see "Data Governance Program Implementation" section below for details). The key individual members and organizational bodies of a data governance program are as follows:

7. Data stakeholders
8. A data governance body
9. Data stewards

Processes: The final component of a comprehensive data governance program addresses all of the processes required for implementing and modifying a data governance program. All policies and procedures for the data governance program should be clearly defined, standardized, and documented. The documentation should outline the "how to" of data governance, including processes required for the implementation of the program; ongoing processes encompassing long-term data management, including measuring program success; and the processes for handling situations jeopardizing data quality or security (e.g., a data breach). The three main types of data governance processes (proactive, ongoing, and reactive) should address, as much as possible, all foreseen data governance activities and describe specific methods for managing data at various stages (e.g., proactive standard setting prior to data collection, ongoing program maintenance, reactive corrections to security policies, and response to a data breach).

10. Proactive, reactive, and ongoing data governance processes

Data Governance Program Components

This section briefly outlines the main focus areas that should be covered by the data governance program.

Decisionmaking authority: Establishing organizational structure with different levels of data governance (e.g., executive, judicial, legislative, administrative, etc.) is a prerequisite to successful data management. This is accomplished by creating a data governance committee, designating data stewards, and defining executive and managing roles and responsibilities at each level of authority (e.g., governance committee members, technology leaders, data stewards, etc.).

Data security and risk management: Ensuring the security of sensitive data (i.e., data that carry the risk for harm[1] from an unauthorized or inadvertent disclosure) and personally identifiable information (PII) by defending against the risks of unauthorized disclosure is a top priority for an effective data governance program. This goal is achieved by establishing a comprehensive data security management plan with a system of checks and controls to mitigate the data security risks. The policies and guidelines should specify rules for work-related and personal use of all organizational computer and data systems, including procedures for data use, assessing data risks to identify vulnerabilities, and handling data security breaches; and explain how compliance with these policies is monitored. It is critical to conduct regular staff trainings and audits to ensure compliance with organizational policies and procedures. The data security and confidentiality plan should be regularly reviewed and modified to stay up-to-date on the latest threats. (For more information on critical security threats to education data, see PTAC's Data Security: Top Threats to Data Protection brief. For details on how education organizations can benefit from the information technology audits, see PTAC's Responding to IT Security Audits: Improving Data Security Practices brief.)

Data inventorying and data content management: Maintaining a complete up-to-date inventory of all records and data systems, including those used to store and process data, enables the organization to target its data security and privacy management efforts to appropriately protect sensitive data. The data records inventory should specify what data elements are collected, provide a justification for their collection, and explain the intended purpose(s) for their use. An organization should regularly review its inventory and revise data management policies to assure that only those data necessary for meeting the justified and documented set of policy, operational, and research needs are collected and maintained. All data elements should be classified by their sensitivity levels (e.g., by evaluating the risk for disclosure of PII; potential for adverse effects for the individual, should the data become compromised; legal requirements to protect the data; etc.) to ensure that appropriate security efforts are applied to protect the data.

Data records management and data access: Ensuring compliance with security policies is accomplished by clearly specifying all activities related to handling data by data stewards as well as users. This includes stating who can access what data, for what purpose, when, and how. A governance plan should provide guidance about the appropriate managerial and user data activities for handling records throughout all stages of the data lifecycle, including acquiring, maintaining, using, and archiving or destroying both regular and secure data records. Additionally, the plan should specify requirements and mechanisms for de-identifying PII data in order to protect individual privacy (e.g., by removing all direct and indirect identifiers from PII data).

Data quality: Identifying strategies for preventing, detecting, and correcting errors and misuses of data is essential to maintaining high-quality data. A proactive approach to data governance requires establishing data quality standards, and regular monitoring and updating of data management strategies to ensure that the data are accurate, relevant, timely, and complete for the purposes they are intended to be used. A robust data governance plan should outline acceptable uses of data that balance privacy and security with the need for high-quality data required for statistical analyses. Periodic quality audits should be built into all cycles of data management, including collection, reporting, and release.

Data sharing and reporting: Ensuring that data dissemination activities comply with federal, state, and local laws is a key organizational responsibility. The release or sharing of any data without written consent (e.g., in the form of individual records or aggregate reports) must adhere to the policies and regulations established by the organization, including procedures for protecting PII when sharing with other agencies and disclosure

[1] Here, harm refers to any adverse effects that would be experienced by an individual whose PII was the subject of a loss of confidentiality, as well as any adverse effects experienced by the organization that maintains the PII (National Institute of Standards and Technology Special Publication 800-122, (Apr. 2010), p. 3-1, 2). Harm to an individual includes any negative or unwanted effects (i.e., that may be socially, physically, or financially damaging).

avoidance procedures for protecting PII from disclosure in public reports. (Make certain that any data sharing agreements are allowed under local, state, and federal privacy laws and regulations, such as the Family Educational Rights and Privacy Act [FERPA]). Further, the data governance plan should specify procedures for regular stakeholder notification about their rights under federal, state, and local laws governing data privacy.

Data Governance Implementation

This section outlines the basic steps for implementing a data governance program and provides best practice advice about each step.

Create a data governance committee: In a complex data system with multiple uses and users, the task of establishing and promulgating data stewardship policies and procedures is assigned to a data governance committee. This committee should be cross-functional and include representatives of management, legal counsel, the data system administrator, data providers, data managers, privacy and security experts, and data users from across the organization. The committee members representing these different offices (e.g., offices involved in contributing or using data) should be appointed by these different areas (or a group of individuals) with executive management authority, such as the head of the state education office, school district, school, etc.

Define data governance policies and procedures: The governance committee should work collaboratively to develop the policies, procedures, and standards for a privacy and data protection program. These policies and procedures should be workable across different levels of the organizational data governance structure. The governance committee is responsible for formalizing the written data governance policies and procedures after soliciting stakeholder feedback and securing support from the executive leadership.

Deploy the data governance program: The specific policies and procedures outlined in the data governance program should be implemented by the data stewards through the ongoing management of data, including collection, processing, storage, maintenance, and use of student records. Any changes to the governance program should be approved by the official(s) with executive authority who initially appointed the data governance committee members.

Monitor and report program progress: Data stewards are responsible for actively monitoring data-related activities for compliance with the established standards and policies and procedures. The data governance committee should track program implementation progress with key metrics (e.g., data quality statistics) and periodically report on the progress to the leadership group and other stakeholders.

Summary

Successful data management requires a proactive approach to addressing stakeholders' needs for high-quality data, while protecting the privacy of individual respondents. To accomplish this, organizations are advised to develop and implement a comprehensive data governance program. A sound governance program will help organizations to improve their decisionmaking and improve efficiency of operations through establishing a coordinated response to common issues, such as data access controls and staff training; standardizing data definitions and processes; and implementing a holistic approach to mitigating data security risks.

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

Data stewardship can be defined as a comprehensive approach to data management to ensure quality, integrity, accessibility, and security of the data.

Data stewards are managers and administrators within an organization who are responsible for implementing data governance policies and standards and maintaining data quality and security.

Education agency or institution refers to any public or private agency or institution to which funds have been made available under any program administered by the Secretary, if the educational institution provides educational services or instruction, or both, to students; or the educational agency is authorized to direct and control public elementary or secondary, or postsecondary educational institutions. For more information, see the Family Educational Rights and Privacy Act regulations, 34 CFR §99.1.

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.

Personally identifiable information (PII) refers to information that can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other information. Additional information on PII is available in the Family Educational Rights and Privacy Act regulations, 34 CFR §99.3.

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.

Additional Resources

- FERPA regulations, U.S. Department of Education: www.ed.gov/policy/gen/reg/ferpa.
- National Institute of Standards and Technology (NIST), NIST SP 800-122, *Guide to Protecting the Confidentiality of Personally Identifiable Information* (2010): http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf.
- Privacy Technical Assistance Center (2011): *Data Governance Checklist,* available at http://ptac.ed.gov/content/checklist-data-governance-dec-2011.
- Privacy Technical Assistance Center (2011): *Data Security: Top Threats to Data Protection,* available at http://ptac.ed.gov/sites/default/files/issue-brief-threats-to-your-data.pdf.
- Privacy Technical Assistance Center (PTAC), U.S. Department of Education: http://ptac.ed.gov.
- Privacy Technical Assistance Center: *Responding to IT Security Audits: Improving Data Security Practices,* available at https://nces.ed.gov/programs/ptac/pdf/issue-brief-responding-to-security-audits.pdf.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics (2011): *SLDS Technical Brief 2: Data Stewardship: Managing Personally Identifiable Information in Electronic Student Education Records* (NCES 2011-602), available at http://nces.ed.gov/pubs2011/2011602.pdf.