Data Governance Checklist

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 https://studentprivacy.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

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 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 the 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? Make certain that any data sharing agreements are allowed under local, state, and federal privacy laws and regulations, such as FERPA.
- ☐ 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 the 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 are 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 *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII),* (NIST Special Publication 800-122, 2010) for more information.

Additional Resources

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

PTAC Issue Brief, *Data Governance and Stewardship* (2011 revised 2015).

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