Best Practices in Data Governance to Enhance Public Trust In and Use of Education Data

Recommendations for Philanthropic Efforts

Sara Ji

Harvard University

September 27, 2024

Data Governance Challenges and How to Improve

Educational data systems face several challenges on both technical and human sides. A few core **technical** challenges include:

Handling Raw Data

One major challenge in handling raw data is the reliance on manual edits, which often lack proper documentation and transparency. A systematic approach using algorithms and automation can enhance the data cleaning process, though it requires investment in both technology and skilled personnel.

Linking Data

Linking data across various state agencies for P20W systems introduces privacy concerns as traditional methods using direct identifiers. Using secure hashing for ID linkage enhances security by making the process irreversible, but it also requires specialized knowledge and investment. Also, the irreversible nature of hashing means that any accidental errors or omissions in the hashed data are permanent, potentially impacting data quality for certain analyses.

Data Masking and Privacy-Accuracy Tradeoff

Masking data effectively requires balancing privacy with data utility. Differential privacy offers a quantifiable approach, allowing stakeholders to set privacy budgets tailored to different groups. While this enhances security, it requires skilled personnel and potential system upgrades to be implemented effectively.

Organizing Data

Data organization often suffers from inefficiencies due to lack of structure, impeding data governance and access. Implementing robust metadata management systems, including data catalogs and documentation, can significantly improve organization. While the scale of investment can vary, larger implementations may require substantial financial and human resources.

Presenting Data for User Experience

Presenting data effectively is a challenge, with many state dashboards either lacking entirely or being not very user-friendly. Data visualizations can make complex information more accessible and intuitive for users, allowing interaction with data in meaningful ways. Employing a user-centered design in data presentations ensures that different user groups can access the information they need in a format that suits them. However, these tools require technical skills to develop and maintain, along with ongoing updates to meet evolving user needs, which requires significant and consistent investment.

On the **human** side, there are also a few challenges:

Stakeholder Engagement

A major challenge in stakeholder engagement is the lack of communication and understanding among diverse groups, often with conflicting needs. To address this, developing stakeholder personas and journey maps can be important. This approach helps identify and map out stakeholder needs, pain points, and pathways to effectively address or accommodate these needs.

Stakeholder Communication

Stakeholder communication often suffers from a lack of transparency and coordination, even when stakeholder needs are understood. To foster better collaboration, dialogic communication strategies are essential. By organizing stakeholder meetings that bring together students, parents, researchers, and practitioners, we can create an environment for open dialogue and shared understanding.

Lessons from Case Studies

California's centralized Cradle-to-Career Data System enables scalability by coordinating multiple state agencies and implementing cloud-based systems, but it faces communication barriers and requires significant investment to maintain. Texas uses a decentralized model through the Tri-Agency Initiative, which promotes collaboration but struggles with coordination and lacks dedicated governance staff, leading to lower stakeholder engagement. Washington offers a balanced approach, with the ERDC managing data governance through a scalable P20W system. However, inconsistent transparency when engaging stakeholders and poor user engagement raise concerns about trust and data misuse.

These case studies demonstrate the trade-offs between governance models: California's centralization drives scalability but at the cost of complexity and communication; Texas's decentralization model leads to weak governance and affects its scalability; Washington's model is in between California and Texas, and while it's a scalable system, it still faces issues with transparency and engagement due to a lack of a formal governing board. These lessons show how governance structures impact the scalability, stakeholder trust, and effectiveness of education data systems.

How to Improve Education Data Governance

Top Priorities

- Invest in modernizing legacy data systems by helping institutions adopt cloud-based solutions, and upgrade data storage and processing capabilities to accommodate growing volumes of data.

- Adopt privacy-preserving technologies like differential privacy to enhance data security. Funding for staff training to implement these technologies is also critical.

- Establish leadership roles like Chief Data Officers (CDOs) or Data Stewards within educational agencies can strengthen data governance structures. Funders can provide grants to create these roles, along with developing policies and procedures for transparent data management and decision-making.

Long-term Initiatives

- Support efforts to improve stakeholder engagement by providing training programs that focus on data literacy, privacy, and effective data usage. Establish platforms where stakeholders, including underrepresented communities, can participate in data governance discussions to ensure their needs are addressed.

- Provide grants for the development of user-friendly data visualization and reporting tools that make educational data more accessible and actionable for all stakeholders.