TRAVELING THROUGH TIME

The Forum Guide to
Longitudinal Data Systems

Effectively Managing LDS Data

National
Forum
on Education
Statistics

TRAVELING THROUGH TIME

The Forum Guide to
Longitudinal Data Systems

Effectively Managing LDS Data

National
Forum
on Education
Statistics

National Cooperative Education Statistics System

The National Center for Education Statistics established the National Cooperative Education Statistics System (Cooperative System) to assist in producing and maintaining comparable and uniform information and data on early childhood education and elementary and secondary education. These data are intended to be useful for policymaking at the federal, state, and local levels.

The National Forum on Education Statistics (the Forum), among other activities, proposes principles of good practice to assist state and local education agencies in meeting this purpose. The Cooperative System and the Forum are supported in these endeavors by resources from the National Center for Education Statistics (NCES).

Publications of the Forum do not undergo the formal review required for products of NCES. The information and opinions published here are those of the Forum and do not necessarily represent the policy or views of the U.S. Department of Education or the National Center for Education Statistics.

February 2011

This publication and other publications of the Forum may be found at the NCES websites.

The NCES World Wide Web Home Page is *http://nces.ed.gov*
The NCES World Wide Web Electronic Catalog is *http://nces.ed.gov/pubsearch*
The Forum World Wide Web Home Page is *http://nces.ed.gov/forum*

This publication was prepared in part under contract number ED-06-CO-0056 with the Council of Chief State School Officers. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

For ordering information on this report, write:

ED Pubs
U.S. Department of Education
P.O. Box 22207
Alexandria, VA 22304

Or call toll free 1–877–4ED–PUBS; or order online at *http://www.edpubs.gov*

Technical Contact:

Tate Gould
202–219–7080
tate.gould@ed.gov

The Longitudinal Data Systems Task Force

Chair

Bruce E. Dacey
Delaware Department of Education
Dover, Delaware

Task Force Members

Nancy Burke
Century Elementary School
Grafton Public School District
Grafton, North Dakota

Corey Chatis
(formerly of the Tennessee Department of Education)
Bill and Melinda Gates Foundation
Seattle, Washington

Laurie Collins
(formerly of the SIF Association)
Mizuni, Inc.
Addison, Texas

Patrick Sherrill
Office of Planning, Evaluation, and Policy Development
U.S. Department of Education
Washington, DC

Nancy J. Smith
(formerly of the National Center for Education Statistics)
DataSmith Solutions, LLC
Vienna, Virginia

Steve Smith
Cambridge Public School District
Cambridge, Massachusetts

Raymond Yeagley
Northwest Evaluation Association
Lake Oswego, Oregon

Consulting Author

Anthony Garofano
Quality Information Partners
Fairfax, Virginia

Project Officer

Tate Gould
National Center for Education Statistics
Washington, DC

Acknowledgments

FOREWORD

The National Forum on Education Statistics (the Forum) is pleased to present *Traveling Through Time: The Forum Guide to Longitudinal Data Systems.* This document, *Book Three of Four: Effectively Managing LDS Data*, is the third installment of this Forum series of guides on longitudinal data systems (LDS). One goal of the Forum is to improve the quality of education data gathered for use by policymakers and program decisionmakers. An approach to furthering this goal has been to pool the collective experiences of Forum members to produce "best practice" guides in areas of high interest to those who collect, maintain, and use data about elementary and secondary education. Developing LDSs is one of those high-interest areas. These systems hold promise for enhancing both the way education agencies use data to serve students and the way they do business, from the policy level to the school office and into the classroom.

LDSs are increasingly becoming the state of the art in education data. These systems move us from relying on blunt, aggregate, snapshot student data; to detailed and timely, student-level data that reflect the student's entire academic history. An LDS makes it possible to not only monitor the success of individual students, but also to identify trends in those students' education records. Freeing educators from guesswork and lessening the burden of painstaking data analysis, these systems provide powerful and timely insight about students and allow educators to tailor instruction to better meet individual needs. An LDS can reveal with great clarity what effects our policies, programs, and decisions have on schools. These systems allow agencies to track students across institutions to facilitate appropriate course placement and to determine who has transferred and who has dropped out. Longitudinal data systems also offer a new level of sophistication at the business level that can streamline operations; improve data quality; and free up valuable resources previously allocated to inefficient data entry, maintenance, and reporting practices.

For these and others reasons, states should continue to introduce, develop, and expand their LDSs. The *Traveling Through Time: The Forum Guide to Longitudinal Data Systems* series is intended to help state and local education agencies meet the many challenges involved in developing robust systems, populating them with quality data, and using this new information to improve the education system. The series will introduce important topics, offer best practices, and direct the reader to additional resources related to the LDS development process. In sum, it is intended to help agencies establish LDSs that will have lasting, far-reaching impact on the education system and on students' lives. For a description of the entire guide series, see appendix A.

Book Three of Four: Effectively Managing LDS Data

LDSs provide more information, but these multitudes of new data must be accurate and trusted, collected once, protected vigilantly, and used frequently. This third book in the guide series discusses organizational issues aimed at moving the project forward and ensuring that the data are of high quality so that users may leverage them with confidence for data-driven decisionmaking.

➤ **The introduction** explains the purpose of book three and describes the conventions used throughout the series.

- ➤ **Chapter 1** introduces the concept of data governance, and discusses the benefits of this structure and process for data management.
- ➤ **Chapter 2** lays out a number of basic steps to establishing effective data governance.
- ➤ **Chapter 3** describes the key groups and individual roles that may comprise an education agency's data governance structure.
- ➤ **Chapter 4** defines "quality" data.
- ➤ **Chapter 5** discusses the various organizational processes that facilitate the creation and maintenance of quality data.
- ➤ **Chapter 6** discusses the value of data standards, and describes the major sources of such standards.
- ➤ **Chapter 7** provides a basic overview of issues and relevant laws regarding the protection of student data.
- ➤ **Chapter 8** addresses the need to secure the LDS data to prevent unauthorized access and tampering.

The appendices include an overview of the four books in this series, references, additional resources, a handout on data governance, data governance supplementary materials, and other relevant Forum and NCES resources.
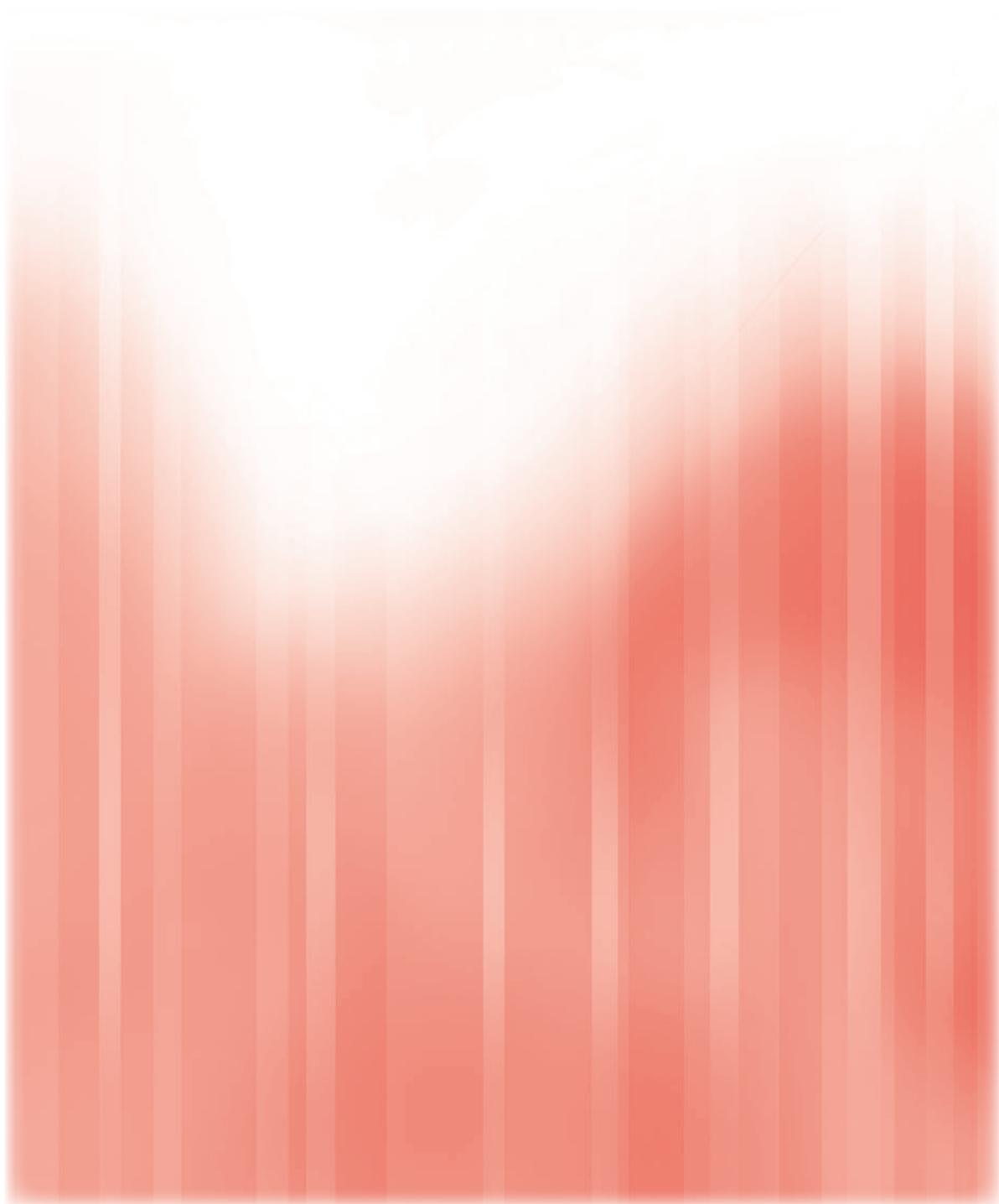
The National Forum on Education Statistics

The work of the Forum is a key aspect of the National Cooperative Education Statistics System (the Cooperative System). The Cooperative System was established to produce and maintain, with the cooperation of the states, comparable and uniform educational information and data that are useful for policymaking at the federal, state, and local levels. To assist in meeting this goal, the National Center for Education Statistics (NCES), within the U.S. Department of Education, established the National Forum on Education Statistics (the Forum) to improve the collection, reporting, and use of elementary and secondary education statistics. The Forum deals with issues in education data policy, sponsors innovations in data collection and reporting, and provides technical assistance to improve state and local data systems.

Development of Forum Products

Members of the Forum establish task forces to develop best-practice guides in data-related areas of interest to federal, state, and local education agencies. They are assisted in this work by NCES, but the content comes from the collective experience of the state and school district task force members who review all products iteratively throughout the development process. Documents prepared, reviewed, and approved by task force members undergo a formal public review. This public review consists of focus groups with representatives of the product's intended audience, review sessions at relevant regional or national conferences, or technical reviews by acknowledged experts in the field. In addition, all draft documents are posted on the Forum website prior to publication so that any interested individuals or organizations can provide feedback. After the task force oversees the integration of public review comments and reviews the document a final time, publications are subject to examination by members of the Forum standing committee sponsoring the project. Finally, the entire Forum (approximately 120 members) reviews and formally votes to approve all documents prior to publication. NCES provides final review and approval prior to publication.

CONTENTS

Introduction

This book, *Effectively Managing LDS Data,* is the third in a four-part series about longitudinal data systems (LDS). The first book in the series, *What is an LDS?,* focused on the fundamental questions of what an LDS is (and what it is not), what steps should be taken to achieve a sound system, what components make up an ideal system, and why such a system is of value in education. The second book, *Planning and Developing an LDS,* focused on the early stages of LDS development to help state and local education agencies determine what they want their LDS to accomplish and what they need to achieve these goals. The present installment discusses organizational issues aimed at moving the project forward; and at ensuring the data are of high quality so that users may leverage them with confidence for data-driven decisionmaking. It looks at the establishment of governance structures and processes, getting the right people in place, and creating committees and working groups of diverse expertise to oversee and inform the process. This process is ultimately aimed at improving data quality and increasing the use of those data to improve education. This book also explores ways to ensure data quality through staff training, validation procedures, and establishment and adherence to data standards. Finally, the document discusses the challenges of securing the system to protect individual privacy and the confidentiality of education records.

Figure 1 lays out the major issues discussed in each of the four books in this series. For more information on the purpose, format, and intended audience groups of this guide series, see *Book One: What is an LDS?*

Figure 1. Summary of the Forum Guide to LDS Series

Book I: What Is an LDS?

- Understanding what an LDS is (and is not)
- Appreciating the organizational steps needed to institute and effectively use an LDS
- Identifying the technical features and capabilities of an effective LDS and the additional features that can enhance the system's utility
- Recognizing the benefits of an LDS

Book II: Planning and Developing an LDS

- Engaging stakeholders
- Describing the current system
- Envisioning the desired system
- Defining needs, including data and functionality
- Gaining buy-in and funding
- Building relationships
- Writing an RFP
- Building or buying a system or components
- Transferring knowledge (e.g., from developers to staff)
- Defining and measuring success
- Refining the system

Book III: Effectively Managing LDS Data

- Defining governance structure
- Defining roles and responsibilities
- Collaborating to improve data quality and streamline operations
- Managing changes to the system
- Training staff to ensure data quality
- Auditing/validating data at all levels
- Establishing/following data standards
- Securing data to protect privacy
- Providing users access to key data

Book IV: Advanced LDS Usage

- Collecting, storing, and delivering key data
- Developing useful reports to fulfill common data requests and needs
- Developing user-friendly data tools to facilitate access and analysis
- Training users to utilize the technology
- Building awareness, understanding, and analytical capacity

LDS Lore:
The ungoverned education agency

The history and consequences

The state education agency's data collection and management practices had come about over time, driven mainly by compliance and funding. Various program areas were created to focus on specific federal surveys, and staff collected the data needed to do their jobs. Program area staff administered the surveys, followed their own quality assurance processes, and maintained and secured the data in their own silo systems. And, of course, data were reported as required by the federal government. Individual managers took their own approach to directing staff and organizing work, and coordination across program areas was limited.

Over time, different departments began to collect some of the same data elements. Inconsistencies were commonplace. For instance, while the student information system listed an "Aileen Hutchinson," who was not in special education, the Special Education Department's system included a girl named "Allie N Hutchinsen." Despite these discrepancies, the staff knew the two records referred to the same student based on other directory information. However, structural differences between the systems—different definitions, formats, and option sets—further complicated matters. For example, Allie's race, White, was coded as "1" in one database and "2" in the other because the systems used different option sets. Furthermore, because program areas defined their own data elements and used different software to manage them, the ability of the agency's many data systems to "talk" to one another varied from limited to nonexistent, burdening staff with redundant data entry work and introducing errors into the system. With no clear requirement for documentation of data processes, methodologies often changed, as happened when Joe left the agency and no one else knew how to produce the dropout rate. The new guy, Steve, calculated it the way he had at his previous agency. No one saw a problem with this, especially since Steve's numbers were lower than Joe's. But when the time came to compare the new rates to previous years,' some staff realized that they were comparing apples to oranges.

Year in and year out, the work was done, but specific tasks weren't assigned to anyone in a consistent manner, and sometimes reports were late or incomplete. And since there was no "official" source for each data element, data for federal reports might come from one source one year, another the following year.

Similarly, without clear guidelines, staff fielded data requests as best they could. When inquiries came in, the recipient would decide to which program area the request should be sent. In one instance, Talia was asked for the school addresses in a district. While McKenna and Vita both managed school directory data in separate systems, Vita was given the task because her desk was near Talia's.

When postsecondary education leaders asked about sharing data, agency officials cried "FERPA!," invoking the federal privacy law because they mistakenly thought it prohibited such exchange (see chapter 7). Security and access protocols were not well understood, and staff often took a lax approach to protecting sensitive information (see "Identity theft in the printer room" in chapter 8). Data quality problems were handled as they arose, but no long-term changes were made to ensure the same issues wouldn't crop up again—like treating symptoms as they flare up, but never working to cure the underlying illness. And since no responsible group or process was in place to identify the sources of the agency's problems, they went undiscovered and data quality issues were just blamed on IT. This went on for some time and, other than the techies, most staff didn't see much of a problem. It was simply business as usual.

The realization

Over time, the country became more interested in education data. Education stakeholders wanted more information for accountability purposes and to better understand what programs and instructional strategies worked. They wanted data to inform decisionmaking at all levels; and to improve administration, instruction, and student performance. The bottom line was, they wanted data from across the agency and they wanted them fast. This changing environment posed many problems for the agency. Requested analyses required linking across silos, or data integration in a central data store. Before this could happen, duplicate, inconsistent data had to be reconciled. However, once the integration work began, more inconsistencies were discovered than anyone had imagined. Data quality had to be a higher priority; security had to be improved; and the data elements collected had to serve business and stakeholder needs, not just meet federal requirements. Better methods of sharing data had to be devised if the agency was to meet the growing demand for a "P–20" system. And better, more consistent protocols were needed to make data sharing more efficient and prevent improper dissemination. The chief information officer decided something had to be done. Having seen a presentation at a national conference, he was convinced a process called "data governance" could help address the agency's problems.

Conventions

Throughout this series, important terms and topics will be highlighted in sidebars. Notable subject matter will be easily identified by the following icons:

Definitions
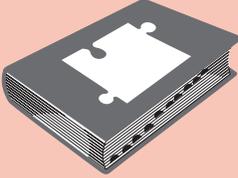
**LDS lore
(fictional tales)**

Key points

Bright ideas

Tips

**District
difference**
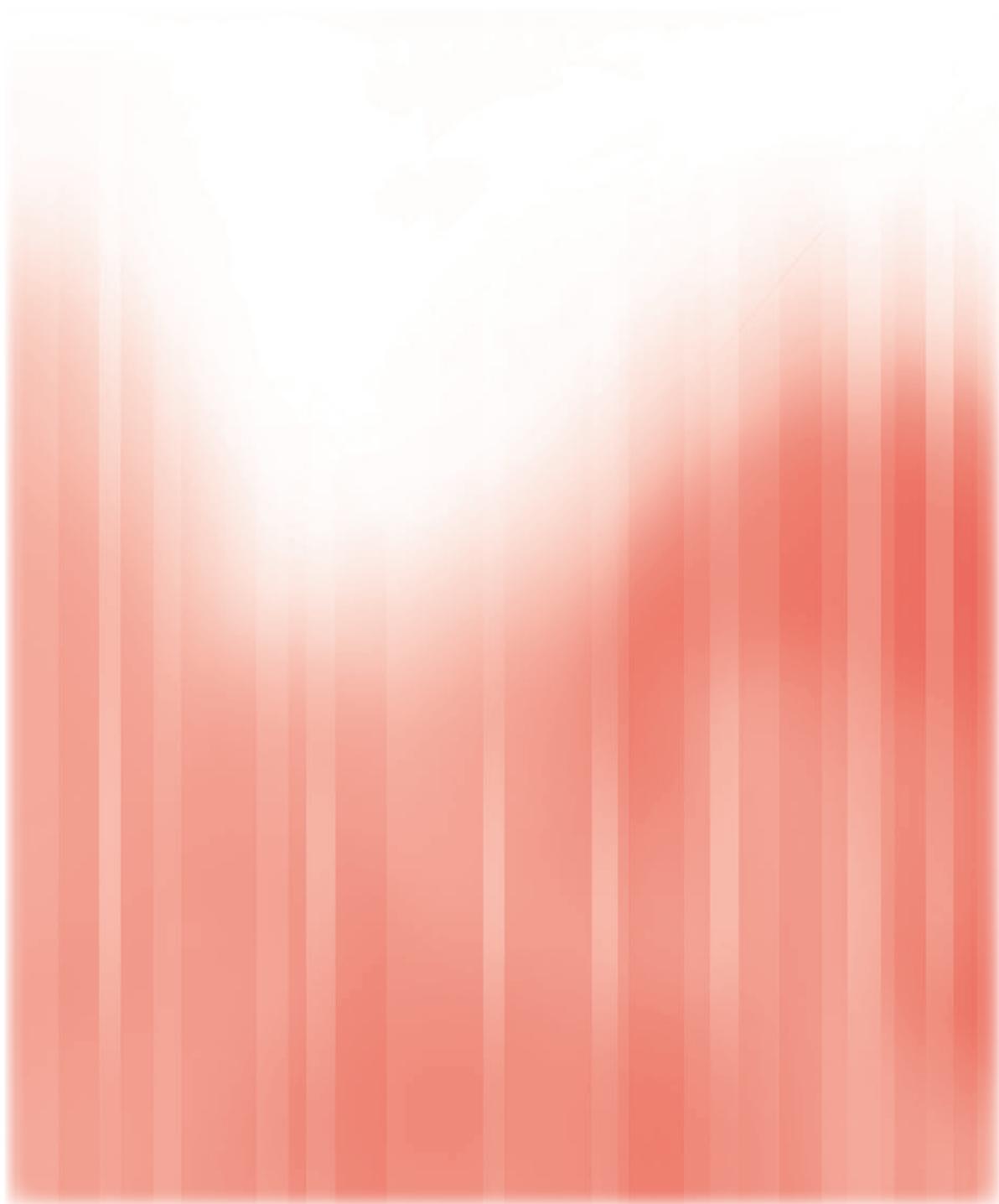
Governing the Data

While certain other sectors have long benefited from good data governance, the education community is relatively new to the concept. Every education agency has some means of handling its data but, historically, many have grown multiple, program area-specific approaches and cultures rather than implement a coherent, well-orchestrated, enterprise-wide system of data governance. Moreover, responsibility for data has too often been unclear or misplaced. For example, data ownership has often been placed with technology staff, who are already responsible for the infrastructure that collects, stores, and shares the data; rather than with program area specialists who have a deeper understanding of the data. This reality, despite its inefficiencies, may have worked well enough when data were only used for compliance purposes. However, the old processes for managing education data must change as the goal of education information systems continues to expand to providing broad access to, and facilitating effective use of, the data. By helping to create greater order, focus, and efficiency, the implementation of a strategic, enterprise-wide system of data governance can help agencies meet their modern goals of data-informed education.

Establishing data governance is crucial to LDS success and should ideally be the first step in the system's development–before implementation begins, before plans are even drawn up. A major effort like LDS implementation expands data quantity and access, and heightens the need for data quality and security. Therefore, a coordinated approach to identifying data issues, creating solutions, and communicating decisions is critical to meeting stakeholder needs. Developing an LDS may allow your agency to collect, maintain, and share data; but without the right policies, processes, and staff in place to ensure quality, these data may be problematic. Data governance is crucial to LDS success, as it creates a culture of accountability, collaboration, and standardization around information. No surprise, then, that LDS development often spurs interest in data governance and provides the opportunity for an agency to spend the time and resources necessary to implement a strong data governance process.

District difference

Data governance at the state and district levels will look very similar. In fact, the information offered in these chapters should be easily transferable to the local level with just one minor and obvious tweak: when involving data suppliers in the process, representatives will come from schools rather than districts.

LDS Lore:
Data governance, ASAP

The agency had planned its ideal system and hired a vendor to build it. Eventually however, staff began to question the quality of the agency's new student-level data being collected and housed. The agency's effort to integrate records from multiple sources was exposing countless inconsistencies and inaccuracies, and it became apparent that potential users would shun the data if they didn't trust them, and rightfully so. Several data-sharing blunders had already occurred, potentially exposing sensitive personal student and teacher information. In addition, while demand for the new data was high and continued to increase, the staff were having trouble keeping up and data request processing was erratic and slow. The LDS investment would be wasted if these problems persisted, and something had to be done to improve data quality and secure the agency's information assets.

Chapter 1

WHAT IS DATA GOVERNANCE AND WHY DOES IT MATTER?

Data governance can be defined as both an organizational process and a structure; it establishes responsibility for data, organizing program area staff to collaboratively and continuously improve data quality through the systematic creation and enforcement of policies, roles, responsibilities, and procedures. As a *structure*, clear and specific roles and responsibilities are assigned and staff are held accountable for the quality of the data they manage. Ultimately though, data governance is not about who
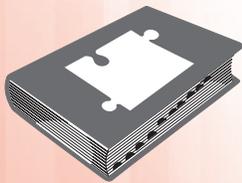
> **Data governance** is both an organizational process and a structure; it establishes responsibility for data, organizing program area staff to collaboratively and continuously improve data quality through the systematic creation and enforcement of policies, roles, responsibilities, and procedures.

is in charge: it is about identifying existing or potential data problems and fixing them to prevent them from happening or recurring. As a continuous and iterative *process*, data governance is a systematic way of handling data throughout the information life cycle, from definition to retirement (see chapter 1 of *Book Two: Planning and Developing an LDS*). The process fosters coordinated responses to ongoing data quality issues and, eventually, a shift to proactive action to stem these problems before they occur. An environment is created in which technical and business data issues can be resolved and prevented in a collaborative, efficient, and transparent fashion. This coordination should extend beyond the compartmentalized program areas and across the business/technology divide, throughout the education agency and even to school districts and other organizations such as postsecondary institutions, the labor department, social services, and others with valuable data related to student histories and outcomes.

Benefits of Data Governance

In an education agency, a data governance initiative typically aims to improve at least three major areas: organizational coordination, data quality, and data use. Figure 2 presents the benefits strong data governance can provide. Operational improvements at the bottom of the chart lead upward to improve the agency's data quality, which then facilitates more effective and widespread use of data to improve education.

Figure 2. Aims and benefits of data governance

Increased use of data to improve education

Higher quality data

- More accurate and reliable data
- More useful data
- Timelier access to data
- More secure data

Greater organizational coordination and collaboration

- Establishment of clear ownership, responsibility, and accountability
- Reduction or elimination of redundant efforts (e.g., data collections)
- More frequent, broader, and better quality communication and collaboration
- Standardization of business processes over time
- Shift in mode of operation from reactive to proactive
- Enhanced understanding of data assets

Greater organizational coordination and collaboration

A more holistic, cooperative approach to handling data can be established through data governance. The decompartmentalization and coordination of enterprise-wide efforts can improve the culture of data collection, maintenance, use, and dissemination. More specifically, data governance accomplishes several goals.

Data governance goes beyond LDS

While data governance is a key factor in LDS success, it provides even broader benefits in terms of how an agency manages data, ensures data quality, and fosters effective use of those data. In effect, any data initiative will benefit from good data governance.

ESTABLISHING CLEAR OWNERSHIP AND RESPONSIBILITIES

As the adage goes, when something is everyone's responsibility, it is no one's responsibility. Data governance assigns responsibility for each and every data element and deliverable to a single data steward, who becomes their "owner." The roles of program area staff are specifically laid out to avoid confusion, and to ensure the necessary work is completed and only one person is accountable for a particular data problem. For example, if discipline data are requested, everyone should know which program area staff member is the discipline data steward and this person, therefore, should respond to the request regardless of who was initially called. In turn, if an issue arises with the discipline data in a legislative report, or on the annual report card, that same staff member is held accountable and is responsible for resolving it. In other words, data stewards are responsible for every aspect of the data they own, from collection to reporting to communication and so on. Thus, data governance helps agencies maintain an orderly operation in which every job is defined and every job is done. Again, responsibility for the data should be seated with program area staff, not the technology staff. The role of IT should be to support the agency's business needs. Responsibility for the contents of those systems–the data–should rest firmly with program area staff.

REDUCING OR ELIMINATING REDUNDANT EFFORTS

With data governance, staff work to identify and eliminate collection redundancies wherever possible. As a result, data elements are shared by all appropriate program areas, but collected only once by a single area rather than multiple times within or across departments. Each element used for federal or state reporting or dissemination has a single authoritative source. Data elements are collected at the individual-record level, rather than in aggregate form, and all aggregate collections end. Ownership of, and responsibility for, all deliverables are clearly documented and communicated, avoiding duplication among multiple staff members. These activities may dramatically reduce the reporting burden of school districts that must report the data, as well as processing time for state staff. Ultimately, data governance will help the agency realize the "collect once, use many times" ideal, improving efficiency and effectiveness as well as the quality of the actual data.

FACILITATING MORE FREQUENT, BROADER, AND BETTER QUALITY COMMUNICATION AND COLLABORATION

Data governance forges lines of communication among a variety of stakeholders. The process is a mechanism for consistent transmission of expectations across program areas within the department, as well as externally with districts and other agencies and partners. Standards for data reporting and collection, and changes to those standards, are clearly broadcast. These include definitions, formats, business rules, responsibilities, and due dates. Beyond facilitating more effective communication, data governance brings affected stakeholders together to collaboratively plan their work and address data-related issues. Program area staff, IT, leadership, districts, and other relevant parties make decisions collectively rather than, for instance, the state agency staff making the decision and simply telling the districts what was decided. Districts and all other affected stakeholders are invited to weigh in on data issues that affect them, and to help create smarter solutions. They are involved in determining new requirements and making sure these can be met, and they also help find ways to make the data more relevant at the local level–a key to ensuring data quality. In short, with data governance, all decisions are made collaboratively and communicated effectively.

IT staff "own" the infrastructure. Program area staff "own" the data.

Data governance helps agencies achieve the "collect once, use many times" ideal.
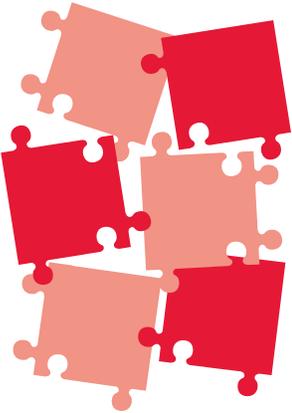
Standardizing business processes over time

Data governance brings staff together to define enterprise-wide standards for each data element, and to normalize the procedures for data reporting and collection. Clearly defined, documented, and well-communicated policies and processes let everyone know what needs to happen, by whom, how, and when. These processes are documented and strictly followed, and each one has a single "owner." For example, while calculating the National Governors Association graduation rate* may involve multiple staff members, only one data steward should be responsible for ensuring that the rate is calculated on time, properly, and in a consistent manner. Guided by clear data management protocols (regarding collection, reporting, sharing, etc.) that are consistent within and across program areas, staff members no longer operate in their own fashion within their own "silos," and the same tasks are no longer performed differently by different people. Everyone, including districts, knows what to expect and what is expected each cycle. Documentation of these processes also helps to ensure sustainability over time despite staff turnover, as well as to increase transparency by detailing processes and clarifying the origins of the data.

LDS Lore: Data quality from management simplicity

More and more work was being done, but the quality of the state agency's data wasn't improving. In fact, the staff's lack of coordination was just making things worse. Cedric and Mark worked together to compile the data for a federal report, running some quality assurance tests on the district data, aggregating them, and building tables. Meanwhile, Amy, the staffer who worked on the prior year's report, had some free time on her hands. Poking around the agency system, she found the new file and took a look. Some of the numbers seemed off, so she ran some code. Her results seemed better than the ones in the file, so she pasted them in...

---

* The National Governors Association (NGA) graduation rate is a standard, four-year adjusted cohort graduation rate agreed upon by all 50 governors in 2005.

Months later, districts started calling to say that some of the numbers in the file were off. Mark and Cedric reviewed and re-ran their code to identify the problem, but they couldn't replicate the reported numbers. No one could figure out what had happened. Eventually, word got around and Amy came forward and admitted what she had done.

The message was clear: Without the assignment of clear roles and responsibilities, and the adherence to clear processes, confusion will result, work will be duplicated, errors will be difficult to trace and resolve, and time will be wasted. In this example, less is more when it comes to data quality, and quantity of work does not necessarily translate into quality of data.

Shifting operations from reactive to proactive mode

Naturally, the data governance process will begin as a reactive course of action that deals with fixing problems that already exist. But, as the pressing problems are solved, the focus will eventually shift to identifying areas that can be improved, and to preventing mistakes from arising in the first place. For instance, if a state report card lists two discrepant counts pulled from two different silos, the agency's reactive response should be to bring the two data stewards together to figure out if there is a reason to use both sources. If not, the next step is to determine which system is the authoritative source for the requested information. A proactive process, on the other hand, would have identified the redundancy before report card season, when everyone worked together to catalog all of the agency's collections and data elements.

Enhancing understanding of the organization's data assets

Through the governance process, staff become more aware of the data the agency collects, and learn which data stewards are in charge of which elements. At a higher level, data governance may help spur a culture shift from viewing data as compliance-driven to viewing them as assets that can help improve their work and student performance. With student-level longitudinal data, agencies can do more than fulfill compliance requirements—they can help improve programs and policies. The data governance process helps stakeholders shape the system to better meet their needs and to expand their capabilities.
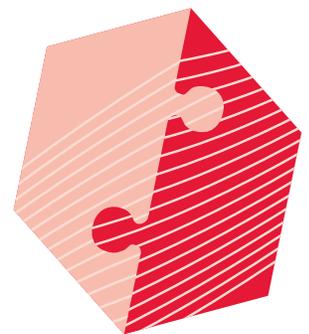
Higher quality data

Implementing a longitudinal data system does not in itself ensure higher data quality. However, it provides an opportunity to improve data quality by bringing errors and inconsistencies to light through the enterprise-wide integration of disparate silos. Without a systematic approach to governing data, the organization has no means of addressing these issues. As a result, the masses of information collected and maintained in the LDS will be questionable and may not meet stakeholder needs. Consequently, many agencies see improving data quality as the primary reason to focus on governance. More specifically, effective data governance will help in several ways.

IMPROVING ACCURACY AND RELIABILITY

Data governance serves to increase alignment among program areas, ensuring consistency in data and related management procedures. Data are more reliable when data managers come together to lay out clear definitions and other standards, and data stewards work to identify and correct any deviations from those standards. In addition, thorough and consistent validation procedures starting at the local level ensure data accuracy; and authoritative data sources are identified and redundancies eliminated, creating a single "truth." These processes ensure that the agency collects data elements only once, streamlining data reporting and making analyses more consistent.

INCREASING DATA USEFULNESS

Data will be more useful when they are aligned with the needs of program areas and other stakeholders, rather than driven by information technology. Once the data governance process helps staff identify, manage, and control all the data, and those data become trusted and reliable, then another level of data governance may manage the use of that information.

PROVIDING TIMELIER ACCESS

Data governance leads to timelier information by increasing the efficiency of data collecting and reporting. Furthermore, districts receive ample notice of changes and, thus, have more time to prepare and work out potential issues. Standard business processes at the state level may also save significant time. Streamlined data-sharing procedures, for instance, can improve how quickly requests are processed, giving users faster access to the information they need. Data governance also serves to standardize records processing from year to year, helping to eliminate time wasted figuring out who will do what and preventing the need to reinvent the wheel every collection cycle.

IMPROVING SECURITY

Within a strong data governance system, staff from program areas and IT work together to determine the sensitivity of each data item and implement effective protections. Clear and consistent data-sharing processes streamline and coordinate agency efforts and help prevent improper release of sensitive data (see chapters 7 and 8).

Increased use of data to improve education

While many see high-quality data as the primary goal of data governance, the ultimate benefit may be their increased use by legislators, administrators, and educators to improve education. For data to be used, they must be accurate and trusted, timely, and designed to meet stakeholder needs. Data governance helps agencies realize these goals by coordinating

Data governance is about people and policies. Technology supports the process but should never drive it.

staff from across the enterprise to collectively solve data issues. Clear data ownership ensures that the right staff members are disseminating and answering questions, and better communication increases all users' understanding of the information collected by the agency. Better quality data equip decisionmakers to improve resource allocation choices, and better student data enable educators to enhance their students' instructional needs. In addition, data governance can culminate in smoother operations and better decisionmaking based on the foundation of timely, secure, high-quality data aligned with agency goals.

These benefits should far outweigh the costs of implementing data governance. If, in fact, your agency has the in-house capacity to implement a data governance process, the cost of a data governance program can be as low as zero. While some state or local agencies will decide to hire an outside consultant to drive the process, others will need to invest only their time. Though changing the way the organization operates and improving data quality will take time, persistence will eventually lead to greater efficiency; time and resource savings; better programs and policies; higher quality information; and, ultimately, better student outcomes.
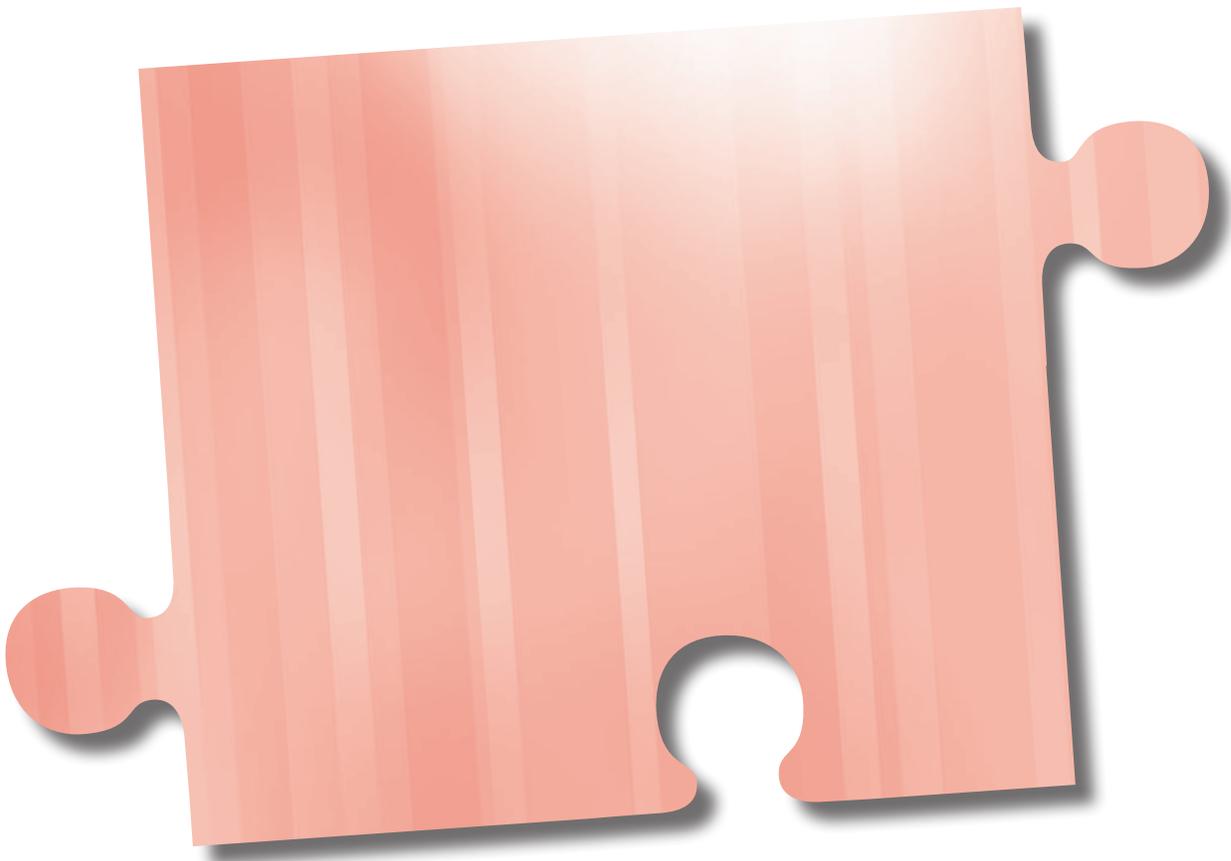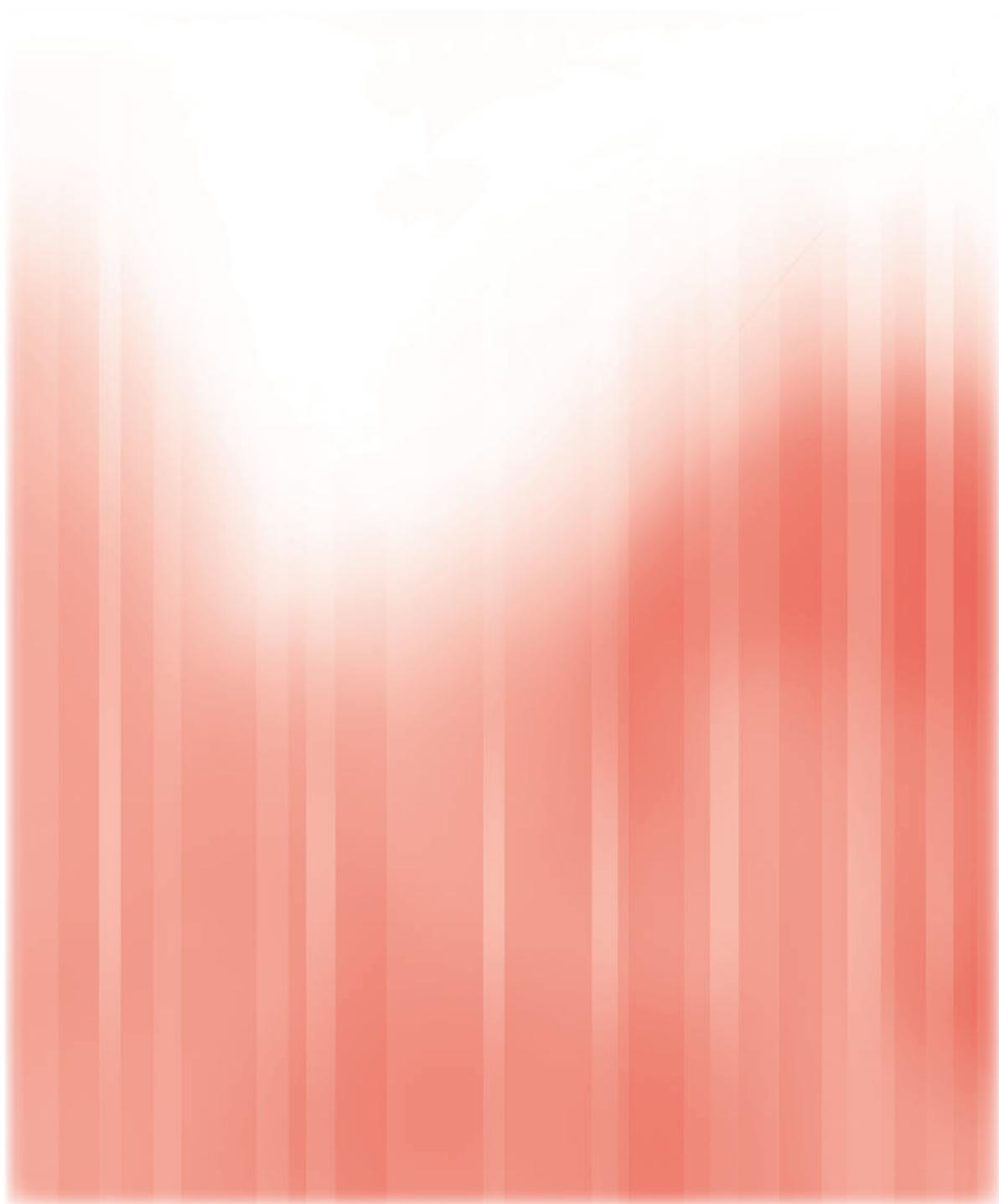
Basic Steps to Establishing Data Governance

Ideally, a good data governance system will be in place before LDS development begins. However, if your organization has broken LDS ground without establishing an effective governance structure, it is never too late to start. And, even if your agency has implemented a governance structure, data governance is an iterative process that can always be improved.

Gauge your governance

How governed is your organization? The *Data Administration Newsletter* offers a data governance test to help you evaluate your agency in eight areas of data governance. The test is available at *http://www.tdan.com/view-articles/10149* and is geared towards building the case for data governance to leaders of your organization.

Although no single approach is best for implementing a data governance initiative, some key action steps should always be taken on the road to good data governance. Normally, chief information officers (CIO) initiate the data governance plan. However they, or other leaders, come to realize their organization would benefit from better data governance, they should begin with a careful and honest appraisal of the agency's approach to data management, weighing the enterprise's current levels of coordination. For tools to help in this period of reflection, see "Gauge your governance" above.

After this stage of self-assessment come some key steps. Firstly, high-level executive support should be sought, followed by the assignment of key responsibilities to staff members and the formation of several groups central to managing and implementing the initiative. Data governance structures range from basic to elaborate in terms of personnel who serve the various functions (see chapter 3). However, the key action steps in a minimal, core data governance structure are:

1. Seek executive support for the initiative.
2. Create a Data Policy Committee.
3. Assign a data governance coordinator.
4. Identify data stewards.
5. Create a Data Governance Committee.
6. Identify, prioritize, and resolve critical data issues.
7. Form working groups of data stewards as needed.

In this chapter, the terms "data steward" and "data owner" are used to describe staff roles within a data governance structure, and do not imply technical or legal ownership. However, these terms are sometimes used elsewhere with meanings different from those applied here. For instance, by "owner," real ownership of the data is not implied here in the way the Family Educational Rights and Privacy Act (FERPA) outlines. That is, at an organizational level, a state education agency (SEA) may be considered a data steward when managing data on behalf of a district, in this case the data owner. Or, similarly, a district might be referred to as a "data steward" for managing data from its schools, which are the "data owners" at this more granular level.

1. Seek executive support for the initiative

Appeal to agency leadership to gain buy-in for the data governance initiative. The specific leaders who should be on board will vary depending on the state's size and staff. In small states, support might come from the education chief; while in bigger states the appropriate executive might be on the chief's staff, perhaps an assistant commissioner or deputy. Communicate the costs of the status quo and the benefits that will result from greater coordination. Executives must understand the value of the agency's data and the need to ensure effective management of those information assets, and to improve quality and security. To gain support, present qualitative and, ideally, quantitative costs of poor data, redundant efforts, and insufficient security. Stress the tangible benefits of a more strategic, enterprise-wide approach to data management that coordinates policies, processes, and architecture to improve data quality; aligns work across the agency and streamlines operations; more effectively protects the data; and shares information in a more systematic and timely manner (NASCIO 2008). In making your case for data governance, try the following compelling arguments.

Fewer errors resulting in lost funding.
Stress real or potential examples of loss of funding due to late or inaccurate reporting to the federal government. Explain how data governance can streamline processes and increase data quality; as well as ensure that the state and districts get the funding they require to meet students' needs. For example, if the special education data were late last year, explain how this might have been avoided with clearer standards and requirements, better communication and collaboration with the districts, improved validation procedures, and the sharing of best practices among data stewards.

More efficient use of resources.

Explain that increased data accuracy and transparency possible with individual-level collections and data governance enable more appropriate allocation of resources. With an individual-level data collection, one can see where the numbers come from rather than just rely on a district tally. For instance, while last year's submission might have included 50 English language learners, this year's student collection may only include 15 individual students. While the aggregate number may be difficult or impossible to verify, the individual records show exactly where the numbers come from.

Saved resources and time through fewer resubmissions, corrections, and audits.

By improving communication between the state and districts, and facilitating collaboration around data issues, data governance reduces the time and money spent fixing bad data (through multiple resubmissions by districts, for example) or auditing districts with problematic data. Better quality data from the start will save time and resources later on.

Economic benefits for the state.

Beyond the effects on the education system, education data may affect sectors of the state's economy. For instance, a mistakenly high number of dropouts may decrease a school's ratings and negatively impact real estate values in the area.

Increased ability to identify common data quality problems among districts, and to target interventions.

Through better coordination with districts, states are able to more easily identify districts experiencing similar difficulties. For example, if different districts use different vendors, the state might determine that a particular vendor needs to make improvements in a certain area. Or, districts might be targeted for professional development or the introduction of new validation procedures.

Fewer headaches in general.

Finally, invoke unpleasant memories, such as phone calls from irate constituents and school staff about bad data. Then explain how these problems could have been avoided had a data governance process been in place.

The importance of executive support for a data governance initiative cannot be overemphasized. Strong and continued commitment from organization leaders will not only provide needed resources but, more importantly, will support the culture change needed in a data governance effort, applying pressure from the top and providing the authority needed to enforce contentious decisions. For instance, a data governance group may make a decision that prompts backlash by a key staff member (perhaps responsibility for a data element is shifted from one staff member to another; or a program area's data are included in the LDS without the department's full support). Without pressure from leadership, staff resistance could undermine the process. Sometimes, just getting staff to show up for data governance meetings can be challenging if leaders do not stress the importance of participation or even make attendance mandatory. If possible, ask leaders to require data governance group members to send qualified substitutes in their stead when they are legitimately unavailable.

The good news is, your agency's leadership is unlikely to willfully resist a data governance plan. Commonly, executives just need to be shown how data governance will benefit the organization. At the very least, staff need to inform leaders of the initiative; ideally, executives will support and participate in the entire process. If leadership support
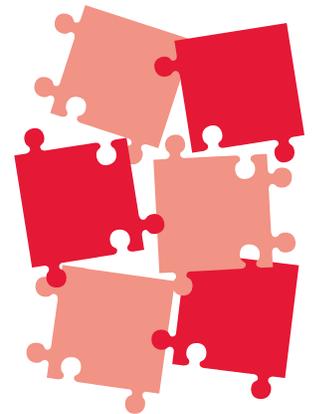
Yori, the data governance coordinator, paced back and forth across the floor of the nearly empty board room. He didn't know why people weren't coming to the agency's second Data Governance Committee meeting. The first one seemed a great success at the time, but here he was with only two thirds of the group, waiting for other members to arrive. He knew that the data governance plan would derail if participation was low. Nevertheless, Yori put on his game face, took roll, and got down to business with those in attendance. After the meeting, he strode down to the education chief's office.

After a compelling appeal, Yori closed the chief's door with a sense of accomplishment. He knew he'd sold her on data governance. The case was not even hard to make. Data governance, from his perspective, was a no-brainer: It was a common sense business solution for improving the department's data. He had laid the case out clearly and simply, contrasting the department's data management status quo with the potential benefits of the data governance initiative. Making the case even easier was the fact that, other than maybe some coffee and donuts for meetings, the costs to the state would be nearly zero.

That afternoon, the chief sent out a brief email:

> "Participation in all data governance meetings is required. Staff assigned to these groups should attend all scheduled meetings, making them a priority over all other activities. If you are unavailable, please send a qualified substitute."

Sure enough, the next meeting of the Data Governance Committee packed the board room. Yori smiled as he called the meeting to order, confident that the plan was back on track.

cannot be quickly won, however, agencies should start or continue their data governance initiatives anyway, while continuing to seek high-level buy-in. Inevitably, this support will prove crucial.

2. Create a Data Policy Committee

The CIO should convene a group of executive management staff that includes the education chief (or other high-level agency staff member), the data governance coordinator, and executive leaders from each program area with a data steward. Rather than creating a new group, identifying an existing group that includes these members and asking

them to occasionally focus on data governance may be preferable. This group's main roles will be to establish the data governance policy; and to address issues that require executive support, such as those that affect multiple program areas and/or affect major agency reports or deliverables. (For a more detailed description of this group, see chapter 3.)

3. Assign a data governance coordinator

Assigning responsibility for overseeing the data governance initiative to one individual is absolutely critical. This staff member, the data governance coordinator, should be the "catalyst" for coordinating the data governance initiative, setting a cohesive action plan, and tirelessly pushing the process forward. Significant culture change must occur for data governance to take hold and make a difference in an agency. In fact, many initiatives fail because no single individual is in charge of making sure that the roles, responsibilities, and processes of the initiative are being followed on a daily basis. In the search for a coordinator, leaders may look for the following:

- ➤ **Tenacity.** This individual must give the data governance process momentum, setting goals and constantly following up on progress to keep the work moving forward.
- ➤ **Strong analytical skills.** The coordinator must identify areas of needs and the players needed to address them.
- ➤ **An ability to see the forest *and* the trees.** This staff member should be able to see the overarching goals of the organization, as well as the details required to meet these goals.
- ➤ **An understanding of technology.** The coordinator should understand technology and be able to bridge the divide between program area staff and IT.
- ➤ **An education background.** This individual must understand why the organization is actually collecting the data and why they matter (it's about the students, not the data).
- ➤ **Good mediation and communication skills.** This staff member must be able to bring people together to work through difficult and sometimes contentious issues.

(For a more detailed description of this role, see chapter 3.)

4. Identify data stewards

Identify areas of data and assign a data steward to each one. Articulate specific responsibilities, making each and every data element the responsibility of a single steward. For example, each data steward should be responsible for data elements, rather than for databases. Data stewards should "own" specific contents of the system regardless of where the data physically reside (e.g., on a desktop, in a database, or in a central LDS). A clear process should be created for identifying data stewards. For instance, leaders may look for staff who

- ➤ have business subject matter expertise and work directly with data (not supervisors);
- ➤ are knowledgeable about data and their educational context, i.e., the programs and policies (preferably not techies);
- ➤ serve as points of contact for districts' questions about a program area's data;
- ➤ are frequent users of data and are comfortable with databases and querying;
- ➤ prepare data for federal and/or state collections;
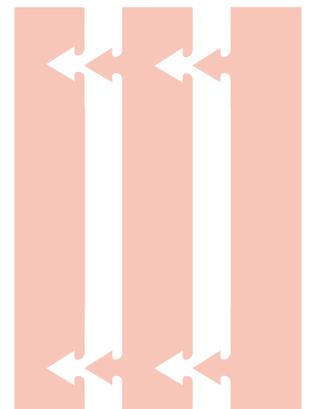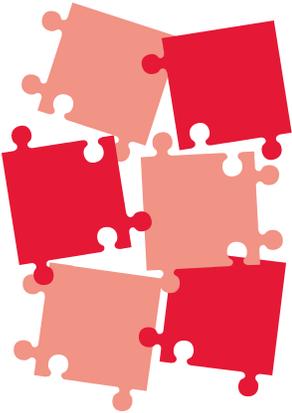
> are detail-oriented and understand how to review data for accuracy; and

> appreciate the value of quality data.

Of course, many education agencies will not have personnel with all of these abilities in every program area. But, as a general practice, rather than hiring new staff to serve as data stewards, start with existing personnel and provide support to grow necessary skill sets and knowledge. Try to identify the best "fit" possible to lessen training requirements. The data governance coordinator should be responsible for identifying gaps in knowledge and skills, and for providing necessary professional development and coaching. (For a more detailed description of this role, see chapter 3.)

5. Create a Data Governance Committee

Form a cross-area group of data management staff that includes the data governance coordinator, data stewards, and other key staff members. This group will be the core of the data governance process, and will effect most of the collaboration and decisionmaking. The data governance coordinator should chair the group and oversee the agenda. Early in its formation, the group should collectively agree upon a mission statement, which should include the identification, documentation, prioritization, and resolution of critical data issues; and to core goals and objectives (see appendix D). The committee should then meet frequently, perhaps monthly, to fulfill its stated mission. The coordinator and the data stewards should contribute agenda items for these meetings. Agenda items may include

> federal reporting updates (e.g., from the ED*Facts* coordinator);

> technology updates, including any IT problems affecting data transmission or reporting;

> LDS project updates; and

> time for any member to raise issues not on the agenda.

(For a more detailed description of this role, see chapter 3.)

6. Identify, prioritize, and resolve critical data issues

Document everything!

Every data governance detail should be documented. For example, who's responsible for what? When are deliverables due? What are the critical data issues and what's the status of their resolutions? What are the standard procedures? Documentation helps keep the work on track, prevents confusion, and allows staff to replicate activities from year to year and in spite of turnover.

At each Data Governance Committee meeting, members should work to identify, prioritize, and resolve critical data issues, maintaining a log to track progress. Critical data issues are the organization's data problems that must be addressed if the committee is to

Master data management

Master data management (MDM) refers to the ongoing process of identifying the authoritative source of data and ensuring that this source is consistently used to feed other data systems, or to populate the agency's central data store; as well as for reporting, dissemination, and analysis. In this way, it is the answer to the "collect once, use many times" challenge. When key data elements are collected and used by multiple data systems, MDM is the process that determines which single source is authoritative. When integrating data from multiple sources into a central data warehouse, for example, authoritative (i.e., "master") sources are identified for each element. And when new elements are collected, an authoritative source is assigned for each. When populating the data store with historical data collected before the MDM process began, the authoritative sources need to be determined for those older data items as well. MDM also keeps track of the data collected and maintained throughout the agency to ensure that common standards (data element names, definitions, codes, formats, etc.) are being used. When all of the agency's past and present data are addressed, MDM's will then focus on handling new data elements.

MDM relies on both data governance processes and technological solutions. The data governance side of the process can be fulfilled through the Data Governance Committee, as it facilitates the collaborative designation of authoritative data sources and elimination of redundant collections. Technology solutions can then be used to share data among multiple data systems ("horizontal integration") by updating secondary data in one system with the authoritative data from another. For instance, if the agency uses several operational databases, the student information system (SIS) may hold the authoritative student addresses, while the transportation system holds secondary student address data. The MDM application would feed the SIS data into the central data store and update the transportation system automatically whenever the information was changed in the SIS.

Additional resources:

■ **The Data Warehousing Institute**

http://tdwi.org

This organization has produced a number of resources about master data management, including a tool to help agencies determine how much a data management solution would benefit them, as well as to assess how ready they are to implement such a solution.

■ **Master Data Management**

http://nces.ed.gov/programs/slds/zip/master_data_management.zip

This presentation on MDM was given by several state education agency representatives at the July 2008 Statewide Longitudinal Data Systems Grantee Meeting.

achieve its core goals. One data steward should be responsible for each critical data issue and, at each monthly meeting, should update the group on progress made towards its resolution. At first, the data governance coordinator will likely identify many of the critical data issues. But as the process matures, data stewards should identify most of the issues. Examples of data issues that might be deemed "critical" are:

> ➤ a data collection that creates significant burden for school districts due to timing, collection mechanism, or duplication with other collections;

> ➤ reporting linked to funding that has been late, incomplete, or inaccurate; and
>
> ➤ high profile reporting that has been late, incomplete, or inaccurate.

Prioritize these issues based on factors such as

> ➤ time sensitivity;
>
> ➤ the number of program areas affected; and
>
> ➤ the data's importance or how often they are used for federal reporting.

(Chatis Consulting)

On occasion, data issues will arise that require leadership-level support beyond the Data Management Committee. The data governance coordinator should bring such cases to the Data Policy Committee.

LDS Lore: Integration-inspired indigestion

Adam dug his fingertips into the armrest. He, his supervisor, and other program area staff were in a meeting with the data governance coordinator and the CIO to discuss the agency's plan to phase its data into the LDS. Yori, the coordinator, was going over the LDS project's goals and timeline, and had just told Adam and his colleagues that their program area's data would be part of the first phase of data migration into the LDS. Adam did not like this idea one bit. "That's my data," he thought. He felt he was losing control of a data set he'd managed for years and he didn't want them dumped into some communal data store. As the meeting continued, he worried the data would be at risk in the LDS and he wasn't swayed by the argument that the move was necessary to give users greater access to the records. "They already get enough access to the data," he thought. Would it be harder for him and his staff to work with the data? He was accustomed to a good deal of autonomy, and he didn't want to have to coordinate with other areas of the department. Suddenly, Adam's stomach tensed. Was this the first step in phasing out, or scaling back, the staff? Would his job be at risk when the integration was complete? Then, twisting his feet around the chair's legs, Adam realized that integrating "his" data into the

LDS would probably expose the records to scrutiny. Errors might be discovered. He and his staff would be held accountable for poor quality data. This was bad…

Across the table, Yori noticed the uncomfortable looks on some of the faces around the table. Adam, slightly contorted in his seat, seemed particularly uneasy. Seeking to reduce the growing tension in the room, Yori reviewed the plan again. He said that, while the integration had support from agency leadership and was nonnegotiable, there would be many benefits for the agency's stakeholders and for the team itself. For example, integration would make it easier to create reports that were previously painstaking or impossible to generate, and conduct analyses across multiple program areas. For instance, they would soon be able to see how their attendance data correlated with discipline incidents and dropouts. Yori said the agency needed the team's help to increase the system's benefits, and asked the staff to consider what reports could be made available through the LDS to ease their workload and help districts. He also assured the team that while the physical location of the data would change, staff ownership of the data would not.

After the meeting, Yori visited each of the data stewards and let them know they should feel free to ask questions or raise concerns whenever they have them. Knowing that some of the staff lacked data management experience, Yori made it clear that this effort was to ensure quality data, not punish people. The agency would offer a host of professional development sessions to inform staff and improve necessary skill sets. Yori also scheduled regular one-on-one meetings with data stewards to see how things were going and to answer questions. During his meeting with Adam, he explained that, while the data governance process was a major departure from how the agency had managed its data in the past, its objectives were aligned with high-level goals and the initiative had strong executive support. Change is difficult, Yori said, but he made it clear they were all in this together and the benefits would be worth the trouble.

7. Form data steward working groups as needed

As discussed earlier, a key principle of data governance is collaboration across the program areas of the organization. Many data issues affect multiple program areas but, without data governance, these issues are unlikely to be addressed in a comprehensive way. According to Chatis Consulting, when an issue arises, the Data Governance Committee should form a working group of stewards to collaboratively address the problems and craft a solution. Within the group, one data steward should be ultimately responsible for ensuring the working group resolves the issue. Groups should identify the problem and pinpoint its original source; define the goals of solving the issue; set up a clear and detailed strategy for a resolution; report back to the Data Governance Committee on progress; work with IT to implement the business solution; and, finally, communicate the resolution to all relevant stakeholders. (See appendix D for guidelines on how group meetings can be structured.)

Track data problems to their source

The identification of a data problem's source is essential if the issue is to be resolved once and for all. If staff only treat the symptoms, the problem will likely surface again in the future.

Teaming Up: The Groups, Roles, and Responsibilities of Data Governance

I n an education agency's data governance system, a range of responsibilities may be assigned to staff to manage the collection, maintenance, reporting, and use of data. This chapter provides a more granular view of the practical workings of the process, detailing the personnel and activities involved; and describing a sample set of groups, roles, and responsibilities your organization may choose to assign in its data governance structure. The information is presented generically and should not be considered the only possible data governance structure. Rather, it is a suggested framework that may be adapted based on your agency's specific needs, staffing capacities, and available resources.

Figure 3 on the following page presents a list of individual roles and groups that may constitute a data governance structure. As depicted in the figure, the data governance coordinator and data stewards work as individuals at the "core" of the data governance process. Several important peripheral roles may initiate, support, inform, or draw from the process.

Several groups should also work to drive the initiative forward, identifying data issues and collectively creating responses. As discussed in chapter 2, the core groups include the Data Governance Committee and the Data Steward Working Groups. Two peripheral groups, the Data Policy Committee and the Data Request Review Board, also serve important functions. The discussion below consolidates many activities into these four groups, though your agency may assign some of these tasks to additional groups.

Individual Role Descriptions

The following draws on the Tennessee Department of Education (draft 2008), DQC 2008, EIMAC 2008, and Rozelle 2006 and 2008.

Several players may be involved in a data governance structure, but a few roles are essential and should always be included. Specifically, every data governance structure should have a data governance coordinator, a group of data stewards, and a Data Management Committee. However, aside from these key functions, many of the roles and responsibilities may be assigned differently than in the manner outlined below.
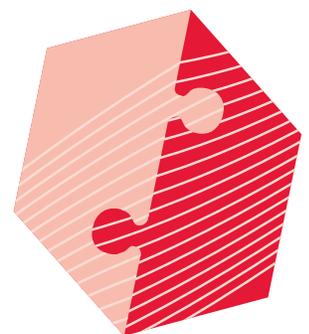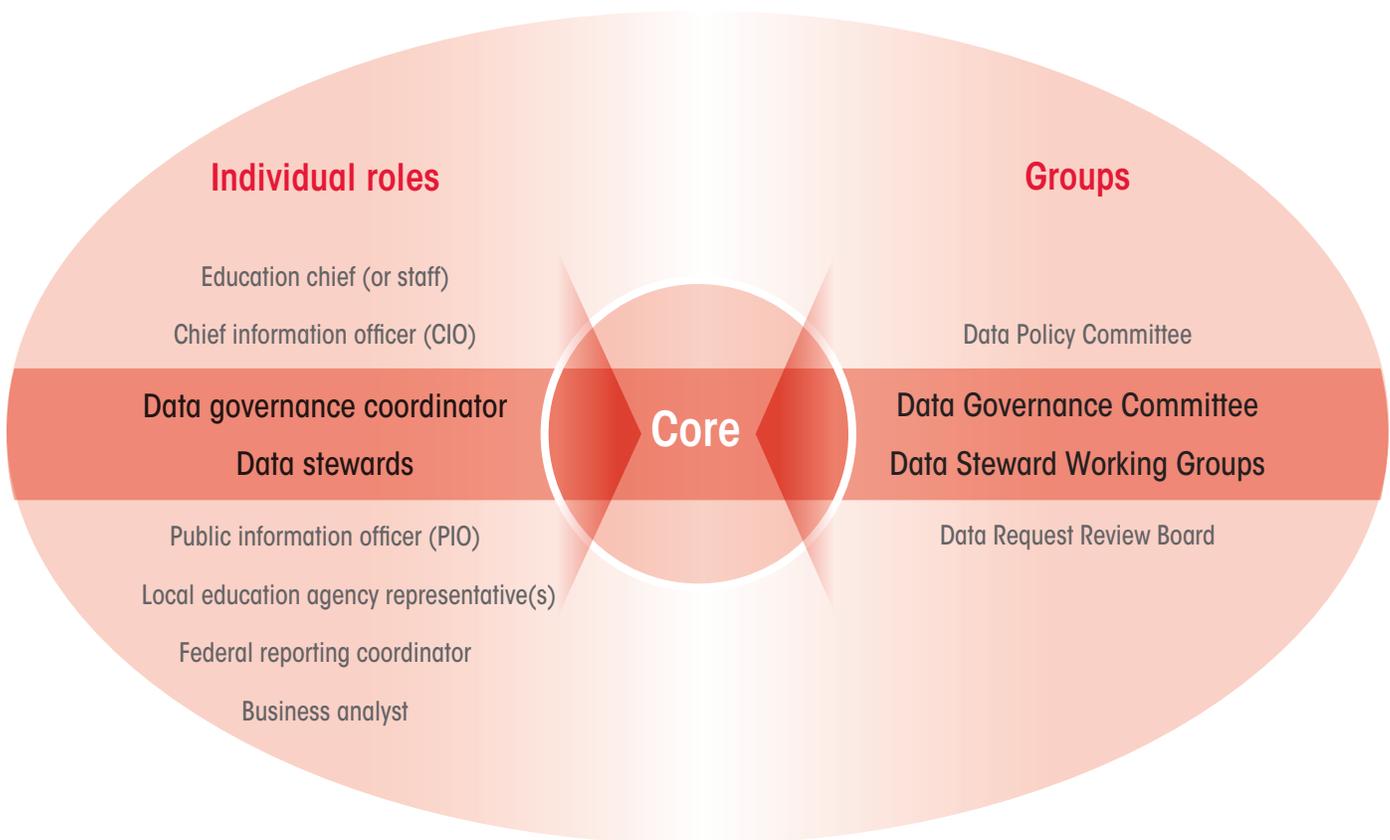
Figure 3. Key groups and roles of data governance

What is important is not necessarily the roles you define, the titles you choose, or how the responsibilities are assigned to each individual. What matters is that all major responsibilities are clearly assigned and staff know their duties. In designating and managing these roles, at least three guidelines should be kept in mind:

➤ Develop and follow clear criteria for selecting staff members for each role. Careful assignment at the onset will help mitigate staff turnover later on.

➤ Provide ample support to help staff successfully meet their responsibilities. Getting everyone up to speed may take time, so persistence and patience are very important.

➤ Reexamine periodically why each role was designated and if each position is filled correctly. Make staffing adjustments if necessary.

Education chief (or other high-ranking staff member)

A high ranking executive such as the education chief, or other high-level executive should provide support for the data governance initiative. Some of this staff member's specific data governance-related responsibilities may include

➤ participating in Data Policy Committee meetings;

➤ mandating staff participation in data governance groups; and

➤ exercising authority to enforce contentious decisions.

Chief information officer

The chief information officer (CIO) usually initiates the data governance process within the organization; and should take the important early steps to establish the initiative and the team. Thereafter, the CIO should remain heavily involved in many of the structure's ongoing activities. Some specific data governance responsibilities may include

➤ seeking executive support for the data governance initiative;

➤ forming, and serving on, the Data Policy Committee and the Data Governance Committee;

➤ working with the Data Policy Committee to develop the data governance policy;

➤ appointing the data governance coordinator;

➤ working with the data governance coordinator to identify the data stewards; and

➤ working with the Data Governance and the Data Policy Committees to identify and resolve critical data issues that require leadership support.

Data governance coordinator

One individual should lead the data governance process: the data governance coordinator, perhaps the most important player in the data governance structure. In fact, no agency should undertake data governance without this single overseer. The coordinator should drive the data governance agenda, direct data steward activities around data quality efforts, and ensure that data issues are resolved. Many data governance efforts fail due to a loss of momentum, and the data governance coordinator must work as a catalyst to keep the initiative moving (NASCIO 2008). This individual must focus on the mission of the group, keeping an eye on the organization's overarching goals, following timelines and managing deliverables, reinforcing the principles of good data stewardship, and working with staff to determine how the agency can better handle its data. The coordinator must stay positive throughout the process because data governance involves significant culture change and this can be hard on staff morale: the constant scrutiny and the focus on improving agency practices can be disheartening. The coordinator must assure everyone that this is necessary and will benefit the agency in the long run. Some specific data governance responsibilities may include

➤ working with the CIO to identify the data stewards, maintain the data steward roster, and train and oversee data stewards' work;

➤ working with the CIO to form the Data Governance Committee;

➤ leading the Data Governance Committee, including scheduling and facilitating meetings, preparing agendas, tracking and following up on action items, and recording and distributing minutes;

➤ participating in the Data Policy Committee meetings by taking data issues to leadership;

➤ convening, and participating in, the Data Steward Working Groups to resolve shared data issues (though individual data stewards should take on greater leadership responsibilities as the data governance process matures).

➤ working with the Data Governance Committee to identify, track, and resolve critical data issues—maintain a log of these issues and ensure a data steward is accountable for resolving each one (see appendix D for a sample critical data issues log);
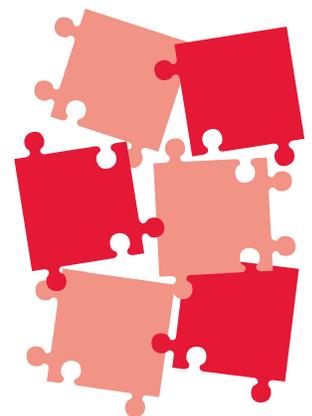
LDS Lore:
Guns and governance

In its Sunday issue, the county newspaper reported that a local school had recorded ten firearm incidents during the previous school year.

The following morning, the school was a ghost town. Frantic parents flooded the office with calls, demanding to know why they had not been notified of these incidents. The staff assured callers that the article was wrong and children could safely return to school: there had been no firearm incidents in the school's history. Zero. A call to the newspaper revealed that the reporter found the information on the state's website. Sure enough, after a few mouse clicks, the school principal located the error. But how could this have happened?

At the state level, the problem was traced back to three possible staff members, but all three said that reviewing those particular data was someone else's job this year. A call to the district found similar confusion: no one could figure out who had dropped the ball there either. Further sleuthing found that a simple typo by an overworked school office staffer had made its way through the quality assurance efforts of both the district and the state, and into the state's LDS. One thing was sure: a pervasive lack of clear roles, responsibilities, and procedures had clearly caused the education community a lot of embarrassment, not to mention unnecessary anxiety at the school.

➤ providing support to data stewards and other staff to ensure all involved know their responsibilities and are capable of meeting them;

➤ facilitating internal communication and collaboration about data quality issues between program areas, and between program areas and technology;

➤ facilitating communication with districts by serving as the main point of contact for data topics; and serving as the district liaison if a state program area undermines or deviates from the data governance policy (e.g., a program area makes a duplicate data request or provides inadequate notice before making changes to a collection);

➤ working with the Data Governance Committee and the IT division to develop and review standards for data elements (see chapter 6);

➤ maintaining a data collection and reporting calendar;

➤ creating and managing the data request approval process to ensure the accuracy and security of shared data (alternatively, the Data Request and Review Board may be responsible for this activity); and

➤ participating in national data conferences and member associations to stay abreast of best practices.

The "data owner"

Some agencies distinguish "data owners" from "data stewards." The "owners" may be program area directors assigned particular, higher level authority for specific sets of data, while much of the work related to managing those data is assigned to the "steward."

Whereas in the simpler data governance structure proposed in this chapter, these responsibilities are assigned to the data stewards, many agencies find this additional level to the hierarchy helpful, even critical. If your agency adopts this extra level of authority, each data owner must support the data governance process and understand that two-way communication with the data stewards is vital. A good rule of thumb is to keep the data governance structure as simple as possible, but whether or not that means adding another level of authority depends on the agency.

Specifically, the data owners' responsibilities may include

- participating in the Data Governance Committee;
- working with the IT division (specifically the security team) to determine the level of security required for each data element they "own"; and
- approving all requests for the data for which they are responsible.

Data stewards (a.k.a. data managers)

Data stewards should be selected to manage the organization's data, with a single steward charged with managing the data and held accountable for ensuring the quality of specific data elements within a particular program area. It is critical that every data element be the responsibility of a single data steward. Note that some agencies divide the steward's role into two levels, including the additional role of data owner in their data governance structure as discussed in the box above (KSDE 2008). Some specific data governance responsibilities may include

➤ participating in the Data Governance Committee (if there is one, this may be the responsibility of the data owner);

➤ identifying and resolving critical data issues involving the data in their charge;

➤ participating in Data Steward Working Groups to collectively resolve data issues;

➤ working with the data governance coordinator to develop and review standards for data elements (see chapter 6);

➤ evaluating data quality (integrity, timeliness, accuracy, and completeness) and management (storage, reporting, adherence to policies, and data architecture);

➤ implementing data quality standards and all decisions of the Data Governance Committee;

➤ ensuring that the data dictionary and the data collection and reporting calendar contain the most current information about their data elements;
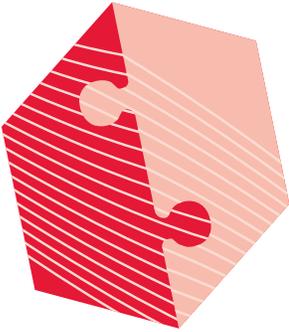
➤ documenting and updating metadata related to data elements for which they are responsible;

➤ working with the IT division (specifically security) to determine the sensitivity of each element and the corresponding level of security and access rights required (while the security team may lead this process, stewards should have substantial input; if data owners are a part of the structure, they may fulfill this task instead of the data stewards);

➤ assisting users with the use and analysis of data;

➤ communicating to districts any changes in the way data will be collected, calculated, or reported;

➤ approving the release of data (for internal or external use);

➤ identifying opportunities to share and re-use data (e.g., for federal or state reporting); and

➤ staying abreast of laws that impact their data.

Public information officer

The public information officer (PIO) identifies and communicates data of interest to the public, and responds to data requests from the press. The PIO should be a member of the Data Governance Committee to stay apprised of data issues, and knowledgeable about ownership and data-sharing procedures. However, PIOs are not data stewards because they are not responsible for any data. When responding to press requests, agency procedures must be followed and data should be released only by the appropriate data steward to reduce the risk of inconsistent or inaccurate information. The PIO may also contribute to the Data Governance Committee by sharing news of any emerging public interest in certain types of data. (This role may also be fulfilled by the communications officer or other staff member with public relations responsibilities.)

Local education agency representative

As they are the source of all data the state agency collects, schools and districts should actively participate in the data governance process to ensure that local perspective informs all decisions. For instance, district representatives will know whether a proposed change to a collection can reasonably be met, and may suggest alterations to facilitate district compliance. Local education agency (LEA) staff can also offer information about district contracts, software, workload, and costs. They can help the state create solutions that will improve its communications and relationships with districts, and effectively decrease the local reporting burden. In addition, participating in the state's data governance process will help increase the districts' awareness of the importance of data quality and provide a model that could be implemented at the local level. At least one LEA representative should sit on the Data Governance Committee, but several would be preferable for a more rounded local perspective (representing districts of various sizes, from different geographic regions, etc.). Some states create an external group of districts and others, such as noneducation state agencies, specifically to review proposed changes to data collections. Some specific data governance responsibilities may include

➤ participating in the Data Governance Committee (as a member, not as a data manager); and

➤ participating in Data Steward Working Groups to resolve data issues that directly involve the relationship and communication between the state agency and districts.

LEA representative selection

State education agencies (SEA) may find it useful to recruit local education agency (LEA) representation in the governance structure to increase perspective. To identify possible candidates, ask state program area leaders to identify district staff who

- contact them frequently about data (both with questions and with suggestions for improving collections—LEA representation will ideally come from both ends of the spectrum);
- care about data accuracy even when the data do not flatter the district; and
- have a program area, rather than a technology, background.

LDS Lore:
The effects of disorder trickle down

The districts were accustomed to supplying the state with a lot of data. Sure, they had to provide the same information multiple times, with some of the elements reported in both student-level and aggregate forms. Sure, changes to data collection requirements were often decreed without prior notice, and confusion about what was required and when it was due. And no, they didn't really get much out of the deal—sending data to the state, but never getting anything useful back. If they realized the data were wrong due to a local or a state mistake, local staff often didn't know whom to contact; and even state staff had trouble directing them to someone who could help. Since the state's system didn't meet local data needs, the district spent its time and money maintaining its own records, while grumbling about the state's shortcomings.

Unbeknownst to the districts, however, the state was implementing a data governance process. Gradually, word got around that several local representatives were working with the state to improve operations. Districts cheered the first sign

of change: an email to the districts listing all of the state's data stewards. Finally, district staff knew whom to call if they had questions about a particular program area or submission! Next came a data collection calendar accompanied by a notice that several of the state's data collections were being eliminated or pared down due to redundancy. District staff gave a collective sigh of relief: not only was some burden being lifted, but they finally had an authoritative source of collection due dates. Of course, data governance hadn't solved every problem, but it was quickly and noticeably improving operations as well as the districts' relationships with the state. At the state level, goodwill and better data from the districts were confirming that their efforts were bearing fruit.

Federal reporting coordinator

The federal reporting coordinator oversees the federal reporting activities across the agency and, as such, should be involved in the data governance process. Some specific data governance responsibilities may include

> ➤ participating in the Data Governance Committee (as a member, not as a data manager);
> ➤ providing periodic updates on federal reporting to the Data Governance Committee; and
> ➤ working with the Data Governance Committee to address issues related to federal data submissions.

Business analysts

In essence, the business analyst resides at the border between the business and technology sides of the agency and, within the organizational structure, may be in either area. Acting as a link between the two domains, the analyst harvests, assembles, and translates business needs into foundational technical specifications. One of the agency's business analysts should actively participate in the data governance process. Some specific data governance responsibilities may include

> ➤ participating in the Data Governance Committee (as a member, not as a data steward);
> ➤ working with Data Steward Working Groups to design the technology component of the solutions developed to resolve data issues; and
> ➤ reviewing data element and technology standards.

Business first, then technology

Program areas should lead the data governance process. Good data governance forces business staff to think more deeply about their needs. If the content area experts don't think through the data's purpose, definitions, and standards, IT will inevitably make assumptions in order to get the job done. If their assumptions are wrong—a common occurrence that is not the fault of IT—data quality will suffer and undue burden could be imposed on school districts. For instance, if a business rule is defined incorrectly (should a student who doesn't earn a high school diploma but enrolls in college be counted as a dropout?) or if the option set for an element does not meet business needs (allowing "Null" in error, for instance), the resulting data may be problematic or even useless. In creating new specifications or amending existing ones, program area staff should work out the business solution to the very last detail, leaving no room for guesswork. Only then should IT be asked to determine how to implement the solution with technology.

Data Governance Groups

The following draws on the Tennessee Department of Education (draft 2008), DQC 2008, Rozelle 2006 and 2008, Chatis Consulting, and KSDE 2007.

In addition to the individual roles involved in the data governance structure, several groups of policymakers, data managers, and other stakeholders should be convened to address data issues collectively. When data issues affect multiple program areas, all affected stakeholders should be included to formulate the best response. These groups also help foster coordination and shared decisionmaking, ensuring that the agency approaches data issues consistently across program areas and over time.

Data Policy Committee

The Data Policy Committee is a group of executive management stakeholders that may include the education chief (or other high-level staff member), the chief information officer, the data governance coordinator, and executive leaders from each program area that has data stewards. Rather than creating a new committee, it would be preferable to identify an existing group that includes these members and insert a standing time slot during regular meetings to focus on data governance issues. The relationship between this group and the Data Governance Committee (see below) is the critical link between leadership and those working directly with the data. The Data Policy Committee provides high-level sponsorship of the initiative as well as leverage for implementing major data-related

decisions affecting multiple program areas. This committee also puts the executive "stamp of approval" on new or amended policies. Additionally, decisions that are contentious or beyond the authority of the Data Governance Committee should be referred to this committee for an authoritative resolution. For example, if staff resists a decision made by the Data Governance Committee, the Data Policy Committee can provide the authority to enforce it. Policies that significantly change the organization's handling of data and data collections should also be discussed by this committee. This relatively small group need not meet as frequently as the Data Governance Committee, perhaps once every other month. Some specific data governance responsibilities may include

> establishing the data governance policy to guide the agency's efforts;

> selecting the data governance coordinator;

> establishing the Data Governance Committee;

> approving data policies and major data-related decisions referred by the Data Governance Committee; and

> identifying critical issues to be resolved by the Data Governance Committee.

Data collection and reporting calendar

Led by the data governance coordinator, the Data Governance Committee should create and maintain a comprehensive, up-to-date calendar of data collection and reporting. The calendar should document and detail all current and planned data collections, including due dates, descriptions of the data and their uses, collection format, and the staff member responsible for each submission. This resource should be made available to all staff and to the public.

Additional resources:

The following resources are available via LDS Share at
http://nces.ed.gov/programs/slds/LDSShare/SLDS.aspx.

■ **Tennessee Data Collection Calendar** (2005)

This is a sample data collection calendar developed by the state of Tennessee.

■ *South Carolina Data Collection Manual* (2006–2007)

This manual lists the data elements collected from the School Administrative SIS by the South Carolina Department of Education. Also included is a table of the various collections, a table describing how the data were used in 2005–06, and a table listing the offices in the department that use the data.

Data Governance Committee

The Data Governance Committee is the center of control in the data governance structure, accountable for the quality of all agency data. It is an enterprise-wide group of data stewards chaired by the data governance coordinator. Ideally, the committee will also include the chief information officer, the public information officer, the lead business analyst, and at least one district representative (though involving multiple districts is preferable). Depending on the scope of the desired data system, this group may also include representatives from other external organizations, such as higher education; and other agencies, such as labor and child services. By including a variety of stakeholders, this group facilitates the collaboration necessary to address shared data issues. Meetings should occur frequently, perhaps monthly or even more if necessary. An effort should be made to make this group act as a resource for data stewards, rather than an obligation. Some specific data governance responsibilities may include

> creating a clear mission statement to guide the group's data governance plan (see appendix D for a sample mission statement);

> working together to identify, prioritize, track, and resolve critical data issues;

> identifying the primary, authoritative sources of each data element (where multiple sources exist), and determining authoritative definitions and values for each data element;

> creating a schedule of deadlines for proposed changes to data collections;

> approving all new data collections and changes to existing collections—a group of district representatives and other affected stakeholders (e.g., other state agencies) should be involved in this process of "change management" (see the "Change management" box later in this chapter);

> creating and maintaining a data collection and reporting calendar;

> eliminating the collection of redundant or unnecessary data—agencies may conduct periodic sunset reviews to evaluate the continued need for every data element it collects; this sizeable task may also be assigned to a separate working group (see the "Change management" box later in this chapter);

> creating Data Steward Working Groups to address shared data issues that affect or involve multiple program areas;

> recommending new policies and policy changes concerning the management, quality, security, and use of data to the Data Policy Committee (via the data governance coordinator);

> overseeing the implementation of the work dictated by data-related policies;

> establishing standard business rules for data collection, sharing, and reporting to streamline operations across the enterprise;

> working with the agency's security team to determine the sensitivity of each element and the corresponding level of security required;

> referring data issues beyond the committee's authority to the Data Policy Committee (via the data governance coordinator);
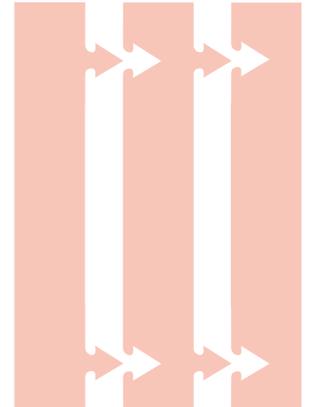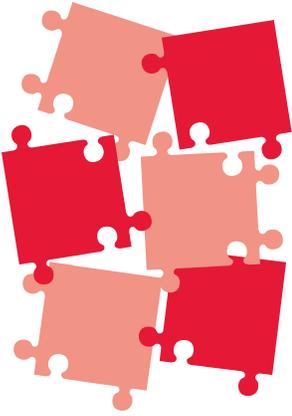
➤ facilitating communication and collaboration about data issues across the agency between data stewards and other staff, and with districts and other external organizations–a standard process for communication should be created and followed, including standard form(s) and frequency of communications; and

➤ reviewing and approving standards for data elements with the IT division.

If the Data Governance Committee does not include much representation from districts and other affected stakeholders, the agency may adopt other strategies for collecting their feedback. For instance, it may hold in-person or online meetings with districts and other affected groups to periodically review proposed data collection changes. It is beneficial to include both program area and technology staff from each agency in these meetings. Alternatively, a separate group of state staff and affected external stakeholders may be formed to take on the responsibility of considering and approving proposed changes.

Data Steward Working Groups

The Data Steward Working Groups are formed by the Data Governance Committee to resolve critical data issues that span more than one program area. Basically teams of data stewards, the groups work together to collectively identify the source of each issue and develop a solution that addresses each program area's needs. While all members of a working group should contribute to formulating solutions, one steward should be responsible for ensuring that the team creates and implements an effective solution. (See appendix D for guidelines for Data Steward Working Group activities.) As Chatis Consulting explains, specific data governance responsibilities may include

➤ defining and documenting the source (not the symptoms) of the shared data issue (communication breakdowns, technology issues, unclear definitions, etc.);

➤ defining the goals of resolving the issue (why the problem should be solved);

➤ creating mini-projects to resolve the problem, including well-documented steps, roles, and due dates;

➤ providing monthly updates to the Data Governance Committee;

➤ notifying the data governance coordinator of issues beyond the Data Governance Committee's authority so they can be referred to the Data Policy Committee;

➤ collaborating with IT to implement the business solution to the issue; and

➤ documenting and communicating the solutions to the Data Governance Committee.

LDS Lore:
The committee becomes a resource

When the agency began its data governance initiative, Gary was tapped as his program area's data steward. He was less than thrilled to find out he'd need to attend a monthly meeting with other stewards and staff from other program areas. At the first meeting, he must have rolled his eyes a dozen times. "What a waste of time," he thought. He had a ton of work to do and needed to call one of the districts as its data were on the verge of being late for the third cycle in a row.

Towards the end of the meeting, Patti, the woman from the English language learner team asked, "Is anyone else having trouble with the district over in Stuckeyville?"

Gary perked up. "Oh, yeah," he said. "The guy Steve over there never returns my calls and they're two days away from their deadline."

"Forget about Steve. Call Fran. She'll help you out," Tara, another data steward, suggested from the other end of the table.

The room erupted as more and more participants began sharing stories and offering tips.

At the next meeting, one of the stewards, Christina, mentioned "master data management."

"Sorry, but what on earth is master data management?" Tara asked.

Most sat silently, a few shrugged; finally, Christina explained what she knew about the subject.

Gary jumped in, "So, it's kind of like horizontal integration?"

"I don't know. What's that?" Steve asked.

Yori, the data governance coordinator, suggested they set up a professional development session. It seemed everyone was at least a little unclear on the subject.

Thanks to conversations like these, Gary and the others started looking forward to the meetings, which they now saw as a resource, and a sense of camaraderie began to form. Staff from previously isolated program areas started helping each other as they realized they shared many of the same problems. The Data Governance Committee had shown them they were not alone.

Change management

While getting a handle on the agency's data early on in the LDS development process is very important, as time goes by, it becomes clear that managing the changes made to the system is also crucial and challenging. As a result, some have developed sophisticated approaches to manage these changes. System changes sometimes impose additional costs and burdens on districts, and introduce data quality challenges. And poor management of change can also be a source of tension between a state agency and its districts. For these reasons, a systematic approach to managing change with the involvement of representatives from school districts and other affected stakeholder groups is essential for the long-term success of the LDS.

A change management strategy should ideally include four main steps:

- Create annual schedules of deadlines for program areas to submit requests for additions or changes to data collections.

- Establish data governance group(s) to review all proposed changes to the agency's data collections. Invite a large and varied group of district representatives to provide feedback on changes. This local perspective will inform better solutions.

- Clearly communicate all changes to districts as early as possible. Effective communication strategies include district staff involvement in the data governance process; periodic training for regional or district staff and school district vendor staff on changes to the data collections such as alterations to record structures, reporting requirements, business rules, standards, edits, etc.); Web meetings; and prominent, public online posting of new documentation such as new standards, due dates, etc.

- Conduct periodic sunset reviews to evaluate the need for every data element the agency collects. Justify the continued collection of each element and record, considering whether the item is required by state or federal law, or is otherwise useful to the enterprise. Eliminating unnecessary data items decreases reporting burdens and increases data collection efficiency.

A lesson from Texas

The Texas Education Agency has been collecting student-level data for nearly two decades. Early on, the agency realized how difficult and costly districts found complying with changes to the state data collection (new data elements, altered codes sets, etc.). The state quickly grasped the need to systematically manage the changes by involving districts and other stakeholders in the process, and effectively communicating all modifications across the enterprise.

Putting its plan into action, the state created a data governance structure that includes two groups of external staff focused on managing changes to the state's data collections. One group, the Information Task Force (ITF), includes representatives from districts, regional education agencies, and other organizations; and receives all proposed changes to the agency's data collections, including sunsets. Approved changes are referred to the second group, the Policy Committee on Public Education Information, which includes representatives from school districts, regional education agencies, and several state agencies (the Office of the Governor, Office of the Speaker of the House, the State Auditor's Office, and others). This committee considers all changes recommended by the ITF and either approves, denies, or suggests alternatives to each proposal.

In addition, every two years, the agency conducts sunset reviews of all of its data elements and records. During these reviews, a group of agency data managers and program area staff evaluate the need for *each* and *every* data item. The findings are presented to the state's data governance committees with, for each item, a justification for continued collection (including descriptions of data usage and any state or federal collection mandates) or a proposal to cease collection. Finally, revised standards (definitions, code sets, etc.) and collection requirements (submissions and resubmissions timelines, record layouts, edits, etc.) are created and posted prominently on a public website. All changes to collections and resulting state reports are also highlighted in periodic training sessions with district, regional, and vendor staff.

This process has been very effective in reducing district burden, and ensuring that districts and their vendors have ample time to prepare for collection changes.

Source: The above is largely based on input from the Texas Education Agency (TEA), May 2009. For more information, visit TEA's Public Education Information Management System website at *http://ritter.tea.state.tx.us/peims.*

Data Request Review Board

A Data Request Review Board may be formed to manage the data-sharing process and to handle data requests, which are likely to increase dramatically once the agency starts collecting student-level longitudinal data. This group of directors, legal counsel staff, and data stewards (or data owners) should meet monthly. Some specific data governance responsibilities may include

➤ creating and enforcing policies and procedures for handling data requests that standardize the review of, and response to, data requests (for instance, standard criteria for approval and denial, such as legality under privacy laws and potential benefits to the educational system);

➤ documenting all approvals and denials of information requests—in addition to internal recordkeeping, this allows staff to more easily identify common data requests that might be fulfilled through a data mart or other "self-service" resource, and it fulfills the recordation requirement under the Family Educational Rights and Privacy Act (FERPA);

➤ prioritizing approved data requests based on factors such as merit and staff capacity;

➤ referring requests to appropriate data stewards, ensuring consistency in data sharing;

➤ monitoring the flow and completion of requests;

➤ referring issues to the Data Governance Committee as needed; and

➤ supporting data-for-fee service—if a request will not benefit the education system, or is frivolous or especially time-consuming (for example, the requestor asks the agency to reformat existing data to meet their needs), the agency may decide to charge a fee to offset the costs of processing the request.

Whether or not the agency forms this group, clear policies and processes for handling data requests should be developed. Establishing standard processes at the onset will increase efficiency and help avoid improper data sharing.

(Kansas State Department of Education 2008)
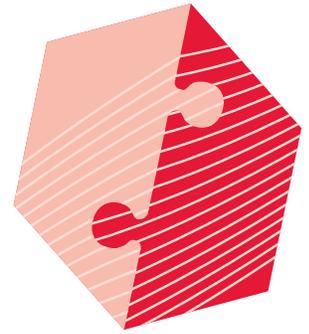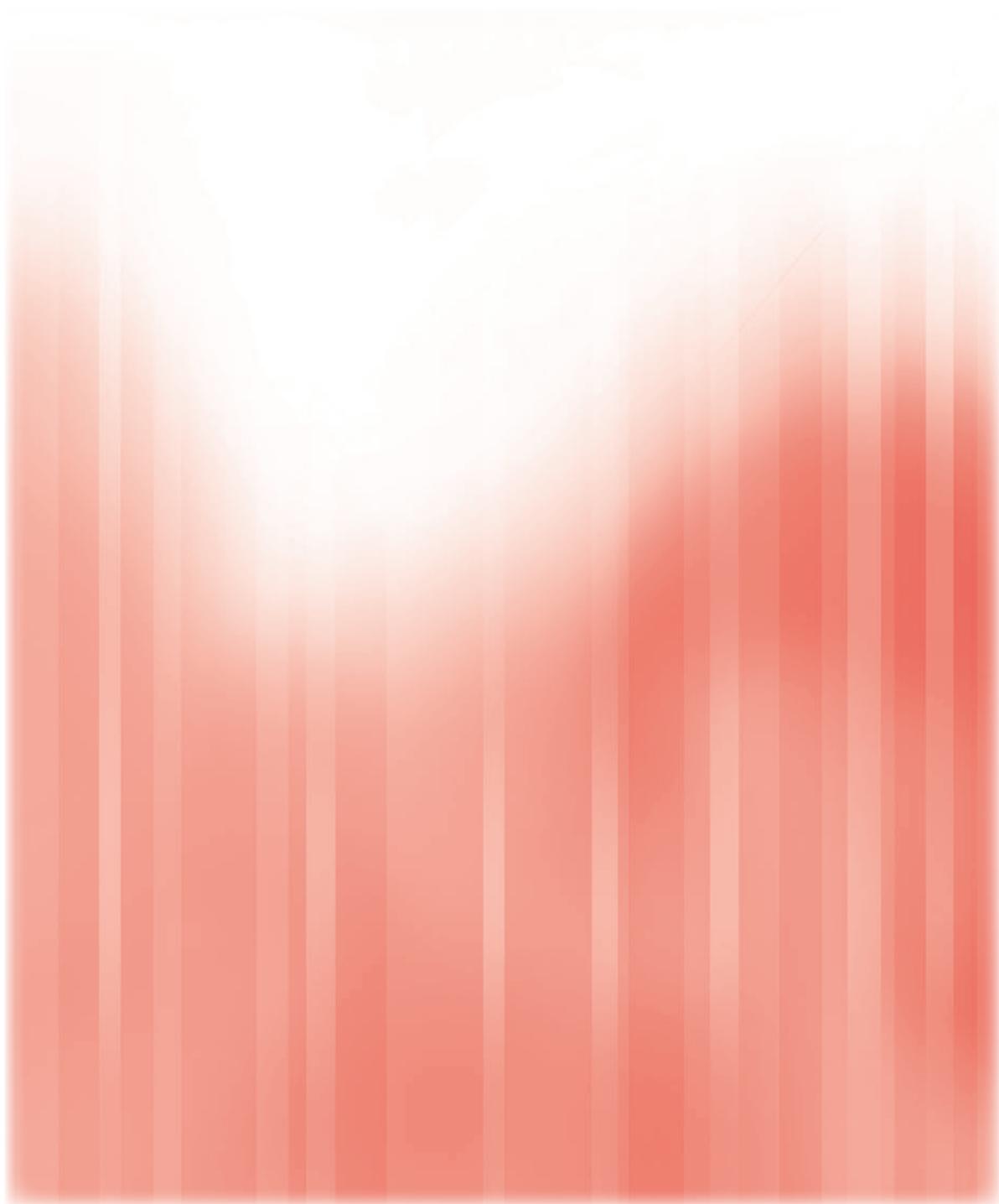
IMPROVING DATA QUALITY

A s education data come to the fore in efforts to improve the education system, so too will problems with those data. The return on your LDS investment is dependent on the quality of the data maintained in, and available through, the system. This reality necessitates a heightened commitment to data quality.

Bad data can lead to bad decisions. In addition, low quality data will not be trusted and, if they are not trusted, the system that maintains them will not be used for better decisionmaking. Worse, inaccurate data can send the wrong message, cause misallocation of resources, or misdirect interventions. Decisions based on misinformation may have potentially dire consequences for individual students, teachers, schools, and districts; and possibly affect funding, reputations, careers, and students' educational opportunities. Fixing bad data ultimately saves staff time and resources. If data are of high quality from the moment they are created, the agency will be able to process and use the information more rapidly and effectively.

Poor quality data come from many sources: data entry and reporting errors, confusion over which data are the "right" data, and inconsistent or ambiguous standards are all common culprits. To avoid costly errors and arm decisionmakers, students, researchers, and other stakeholders with timely, high-quality information, education enterprises must strengthen strategies for creating and managing data. Data quality should be a high priority throughout the agency, with improvement efforts including data governance, clear and enforced policies and standards, careful and competent data entry, quality assurance procedures at all levels, and staff training and professional development. Staff must not only understand agency data procedures and requirements, they must also be convinced of the data's importance. The records must not be seen only as chores; they should be treated as assets that can inform and enhance their work. To this end, it is very important that local staff are able to use the data they create.

The following three chapters aim to help agencies improve the quality of the data they create, collect, store, and make available through their LDS. They provide an overview of many factors involved in creating quality data, and will direct readers to other resources focused on these issues.
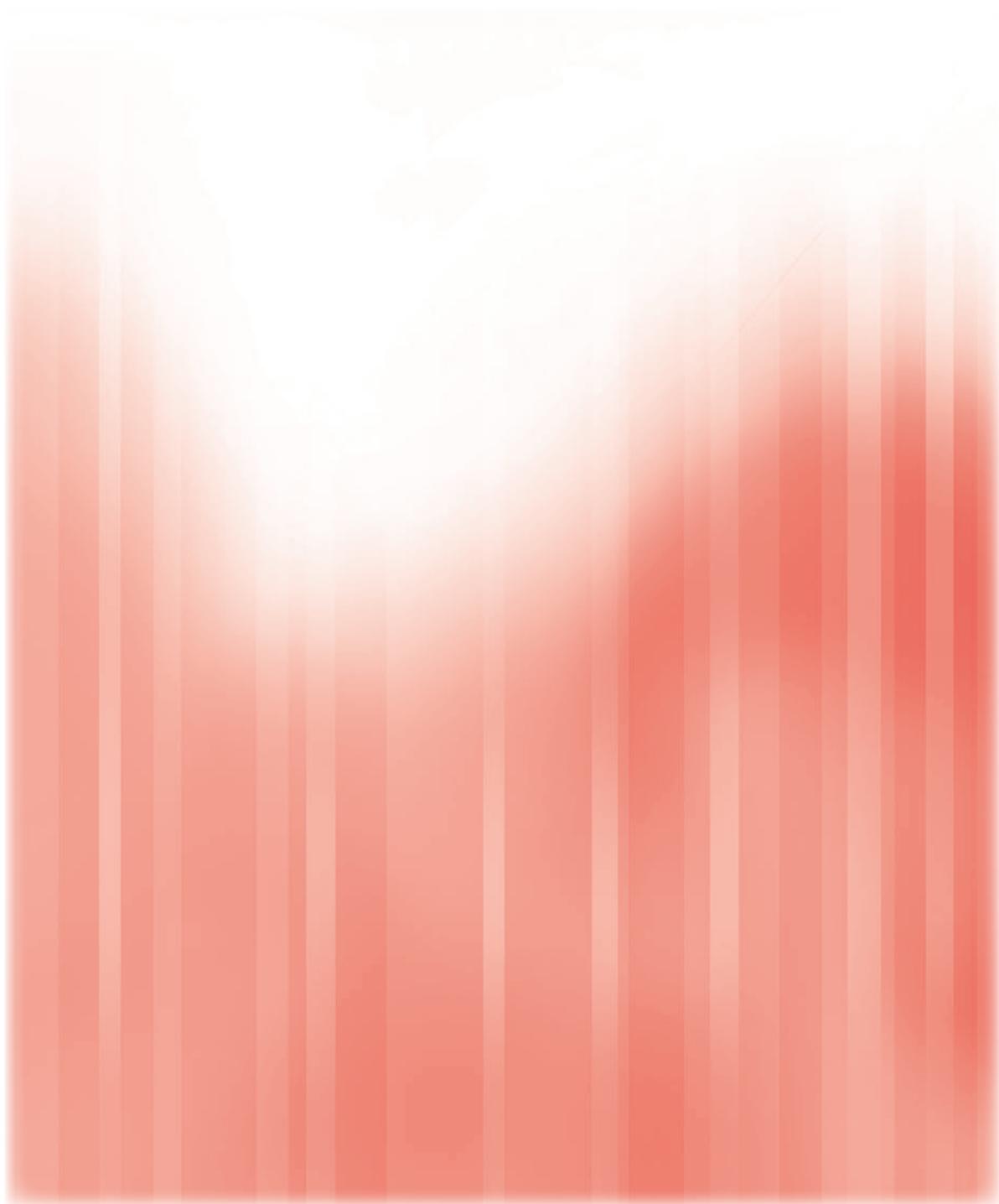
What Are Quality Data?

Quality data are accurate depictions of the real world that are consistent across an enterprise, secure and accessible, delivered in a timely manner, and suitable for their intended applications (Redman 2001).

Accuracy and Completeness

The most obvious measure of data quality is accuracy, or the degree to which the information "accurately" depicts the real world construct or phenomenon it represents. For example, a student's numeric grade for algebra I in the LDS must match the one printed on his or her report card. To be accurate, data reported and maintained must also be complete; for instance, every student must be identified as either male or female.

Internal Consistency

Quality data must be consistent across the enterprise. For instance, a student's name should be recorded in the same manner in every silo system. While a particular child may answer to "Charles," "Charlie," or "Chuck," only one form of his first name should be maintained by the agency. In addition, calculated data items such as the dropout rate should be computed the same way if they are calculated more than once. Internal consistency may also be referred to as "integrity," which may be compromised when data are somehow corrupted during a data transfer or other process. The concept of consistency is also similar to the that of "reliability," which may be diminished if, for instance, the definition of a data element is unclear, leaving room for varying interpretations by the staff creating the data. Additionally, data must be coded in a manner that adheres to defined code sets. Ultimately, inconsistent data will not be comparable for analysis; nor will they be easily interoperable or portable.

Dimensions of data quality:
- Accuracy
- Completeness
- Consistency
- Utility/validity
- Timeliness
- Security
- Accessibility

Timeliness

Quality data must be delivered within a useful timeframe. While a data system may be able to provide teachers with accurate student test scores, the data will be of limited use if they take months to deliver. Thus, while data may be considered of high quality by other measures, they must be accessible to users quickly if they are to meet their intended purpose of providing actionable information for decisionmaking.
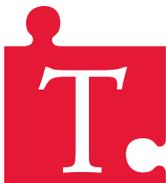
Security Yet Accessibility

Quality data must be secured to protect privacy and to prevent tampering (see chapter 8). At the same time, these data must be available to authorized users to provide information and improve decisionmaking.
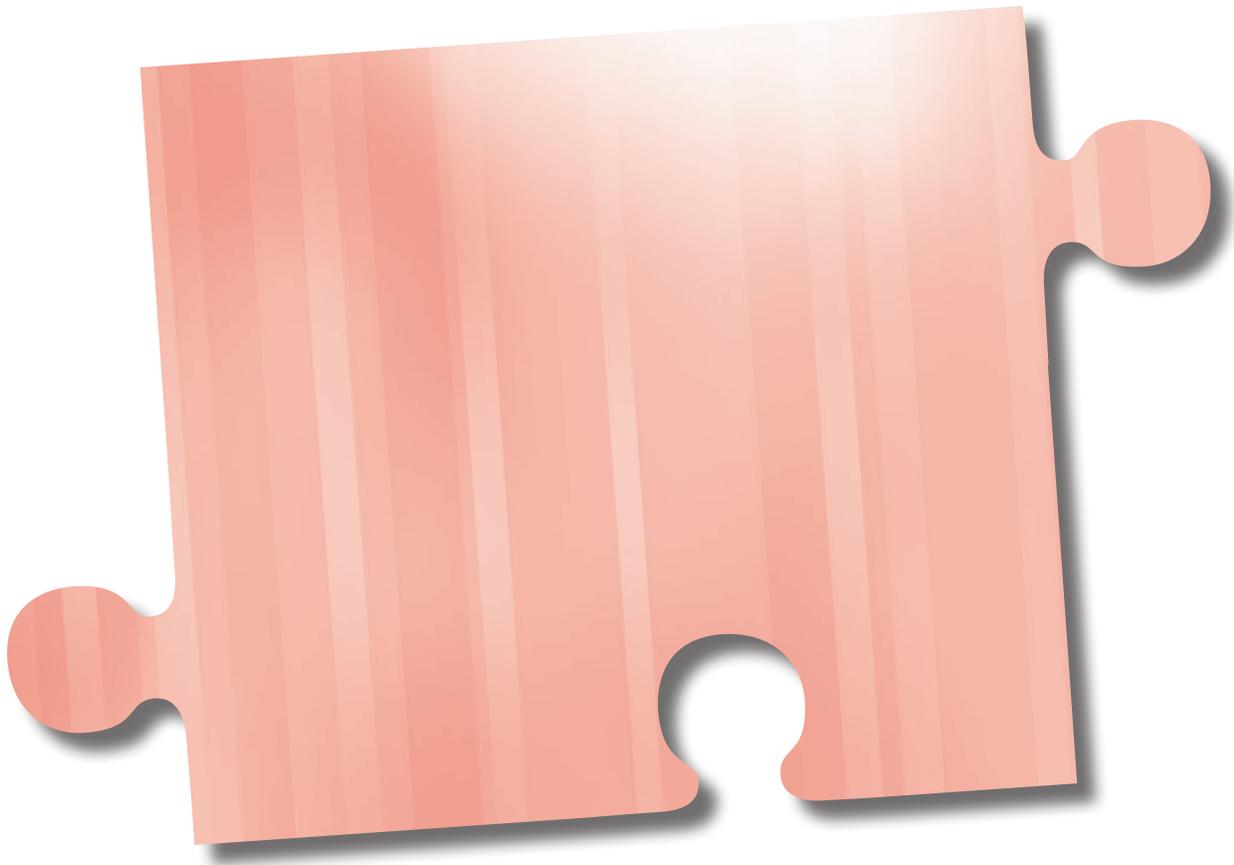
Chapter 5

Data Quality From Bottom to Top

The best way to ensure the quality of data is to get them right in the first place, and to prioritize quality throughout the information life cycle. This approach relies on staff in the school as well as at the district and state levels. It starts at the source, typically in the school where teachers, clerks, and other personnel enter data. From the school, the data are sent to the district, where they are validated and/or audited; then to the state agency and federal government, where further quality assurance processes take place.

The Forum has more detailed information…

…about improving data quality:

- Forum data quality online courses
  http://www.academypa.org/sifa/splashdq.html

- *Forum Curriculum for Improving Education Data:*
  A Resource for Local Education Agencies (2007)
  http://nces.ed.gov/forum/pub_2007808.asp

- *Forum Guide to Building a Culture of Quality Data:*
  A School and District Resource (2004)
  http://nces.ed.gov/forum/pub_2005801.asp

In addition to the processes that check the flow of data "up" the ladder from the school, quality also relies on effective governance and communication back "down" again. Establishing effective data governance at the state and district levels provides a mechanism to help resolve problems and prevent finger-pointing or issue avoidance. Education agencies must move from ad hoc data management models to strategies that bring together all stakeholders from across the enterprise, create key governance groups, assign clearly defined roles and responsibilities, secure agency data, and ensure the data help achieve organizational goals. (For more information on data governance, see chapters 1–3.)

State agencies and the federal government also establish policies, guidelines, standards, and reporting requirements that must be effectively communicated "down" to the data suppliers to enable successful and timely implementation at the local level. Likewise, school districts may create their own standards, guidelines, policies, and regulations to guide school data activities. They create data reporting calendars, data dictionaries, metadata systems, and business rules; assign responsibilities; and implement technology to facilitate data processes. These guidelines and procedures should be similar across program areas so that schools will have comparable experiences submitting various data. In other words, submitting data to one program area should not be very different from submitting to another.

Responsibility for data quality should ultimately rest with program area staff, rather than information technology staff. While IT can perform basic checks to see if the data "look" right, program area staff have a deeper understanding of the information and are better equipped to find errors. For instance, a data report might look right to IT if a total number of schools is generated, but a program area staffer may know if that number is actually correct. This in no way means that IT staff and the technology they manage are not critical factors in improving data quality. Technology that streamlines and automates data entry and sharing is indispensable to quality, as are validation procedures implemented through technology. However, to ensure data quality, education agencies often focus too much on technology and not enough on the data or the people and business processes regulating them. If reported data are inaccurate from the start, the best technology solutions will fail to transform them into quality data.

While the state education agency can do several things to improve the quality of its data, it is essentially a data "receiver," relying on school and district staff to provide quality information.

Data quality certification

The Kansas State Department of Education has been a leader in improving data quality at the local level. The state has created a professional development program that trains and certifies a range of school and district staff on data quality practices and techniques, as well as software applications. For more information on Kansas's Data Quality Certification program, visit *http://www.ksde.org/dqcprogram.*

Provide Training and Professional Development

At the local level, where the life cycle of information begins, the data "creators"—the school teacher, counselor, nurse or secretary entering student data to the district or regional service agency; or vendor staff member building a report for the state agency—must be trained to ensure they produce high-quality data. This training should include best practices and procedures for creating and entering data; and the use of the technology employed to collect, edit, and report data. Staff should also be very familiar with all relevant policies, data standards, reporting requirements, and timelines.

Data appreciation leads to better data quality

Staff preparation should teach more than policies and procedures. Professional development programs, and ongoing communications throughout the enterprise, should help everyone understand why data are so important. Staff need to know how their handling of data affects the use of those data at all levels, shaping decisions from school funding to individual student learning. They must understand why the data are collected, how teachers and decisionmakers use the information, and how the work relates to the money the school or district receives. Understanding their uses will help staff appreciate why data must be accurate and timely, and provide an incentive to strictly adhere to procedures.

Data quality results from data use

If staff see data collection and reporting simply as chores to perform for an authority, they may not be sufficiently motivated to go the extra mile to ensure quality. To create an incentive to improve data quality, agencies must ensure data are used down to the school office and classroom levels. For instance, data will be more useful to practitioners if they have access to student-level data with reporting and analysis tools, or dashboards. District administrators can access the data to see how their schools compare with similar districts in the state. Teachers can view data in real time to inform lesson plans and tailor instructional strategies. Additionally, state agencies may turn the submitted data into useful reports for schools and districts perhaps enrollment lists or comparisons with other schools and districts. If data creators see that the data are used for high-stakes calculations to make their jobs easier, or to hold them accountable, they will have greater incentives to ensure the data are of high quality.

Data Auditing Procedures

The flow of data from schools and districts to the state LDS should include several safeguards to ensure quality. For example, on the way from the school secretary to the district and on to the state data system, certain procedures and mechanisms should be in place to check the data's quality, identifying and resolving anomalies. Ideally, data should be checked for quality before they are loaded into the collecting systems. Some states use auditing mechanisms to check submitted data for problems, and validation reports to alert staff if they find any. Audits may include, but are not limited to, the application of business rules that
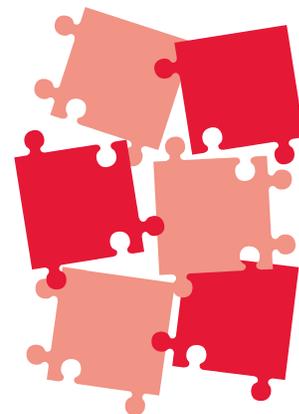
- compare data to prior year values to spot any significant changes where little or no change is expected (e.g., a change in a student's race);
- identify invalid values (e.g., "Null" in a field that requires a numeric value, invalid codes, incomplete or blank fields, out-of-range or over-limit values);
- identify invalid formats (e.g., a date entered in incorrect format); and
- detect excessive use of certain codes (e.g., the frequent use of "Other").

Such front-end validation procedures are preferable to back-end cleansing of data already loaded into the system. In addition, errors should ideally be corrected in the source files and resubmitted, rather than amended in the state or district's system. If a correction must be made by state or district staff, a process should be in place to ensure that the source files are also corrected (Schutte et al. 2009). Errors may be identified as critical, which would require correction; or noncritical, which might require staff review, but not necessarily correction. Data suppliers should be notified of these discrepancies and be required to correct errors or confirm that any questionable data are, in fact, correct. Agencies must determine how often their data will be validated, and establish timelines for submissions and resubmissions. To further validate the information, some states also verify reports with district program area staff before the data are finally released to ensure that the numbers match districts' expectations. This type of process helps guarantee the data reported are accurate representations of the reality.
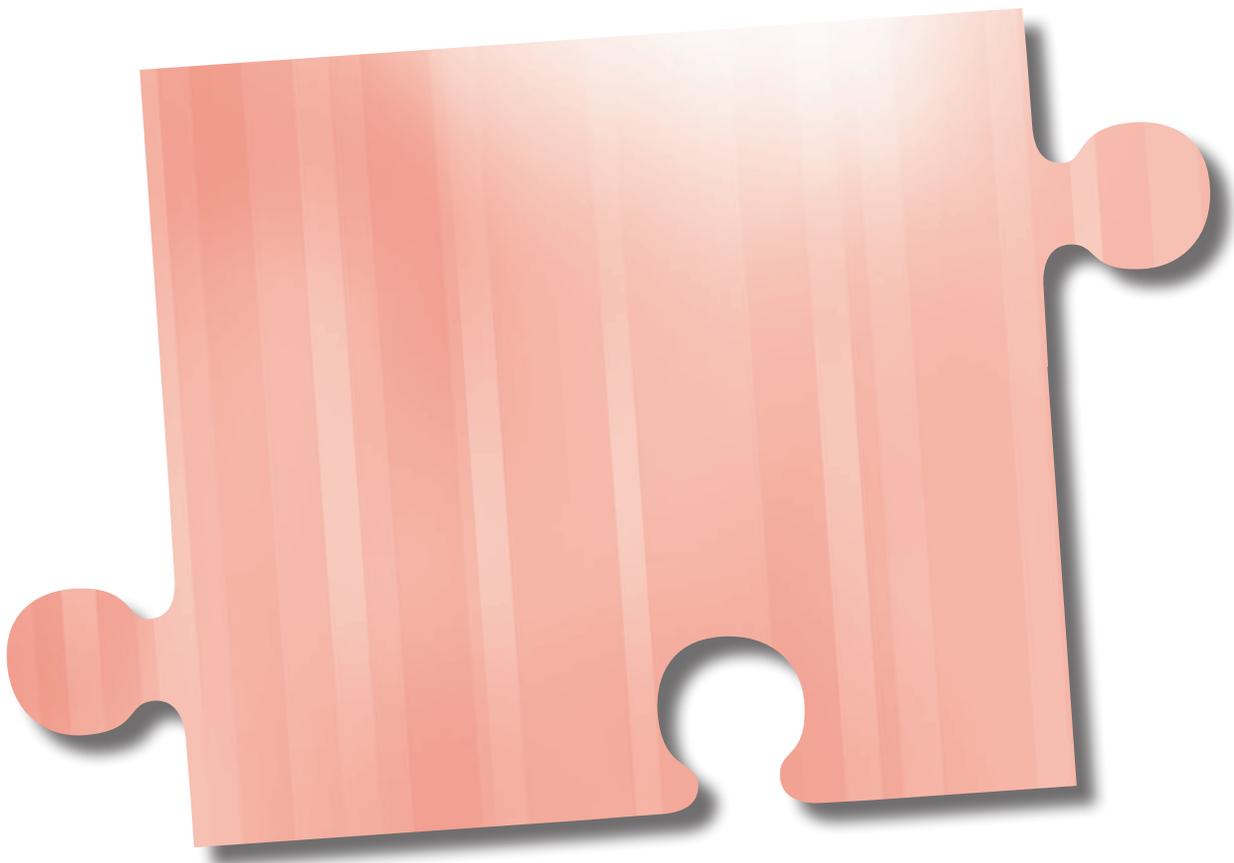
Chapter 6

Speaking the Same Language: Data "Standards"
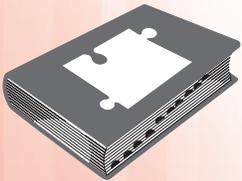
Data standards are documented agreements on representations, formats, and definitions of common data; and are intended to improve the quality and share-ability of education data.

A look around the country reveals that many states and districts are building separate and dissimilar LDSs. While many see this as problematic, potentially complicating the exchange and comparison of data in the systems, trouble actually only arises when the different systems use incompatible standards. Trouble-free data sharing within an agency and with external systems will largely depend on the standardization of data elements and technical specifications across systems, as will the quality, comparability, and usefulness of the data. Adherence to common data standards is the key to bridging these systems, achieving interoperability, and enabling analysis across institutions.

Data standards are documented agreements on representations, formats, and definitions of common data elements; and are intended to improve the quality and share-ability of education data. Under the umbrella of data standards, three major types of standards serve disparate audiences and purposes:

➤ **Data definitions and code sets**

Definitions and code sets are concerned with the meaning and content (e.g., values) of data elements. In a sense, these standards provide a common vocabulary or language for anyone who manages or uses education data. Sometimes called "suggested standards," they are typically included in data dictionaries, glossaries, and various other resources. These standards may be

useful to a broad array of users, including teachers, district and state data staff, researchers, institutions of higher education, and private sector organizations.

➤ **TECHNICAL SPECIFICATIONS**

Technical specifications are used by software and systems developers to facilitate interoperability between applications or to guide data reporting. They typically provide technical criteria or requirements, methods, and processes for data reporting and management such as field length or data type. Technical standards are useful to software application vendors and system developers, as well as education agency staff who submit data to a collecting agency.

LDS Lore:
What's a school anyway?

Unexpected difficulties may arise during the development of standards. For instance, one state agency spent months defining "school," a task that no one had anticipated would be particularly difficult. Staff wrestled with the nuances. Should special education schools be included? How about private schools? Could two schools reside in a single building or one school span across multiple buildings? The choices were anything but obvious. Variations in definitions like these can have serious implications, including on funding allocations.

➤ **GENERAL GUIDELINES/RELATIONSHIPS**

General standards describe the relationships between data elements. Data models are the typical source of such information. These models are commonly developed for software and systems developers to help build data system architecture, as well as for researchers who need to explore the types of data available to them for study.

Use Common Definitions and Codes

Adherence to common data definitions and codes facilitates comparability, interoperability, and portability within and among K–12 institutions; as well as with early education, postsecondary, and workforce organizations. Without a standard set of common data elements, there would be no way to make sense of the data collected and shared across schools and districts, or to truly follow students over time as they change

schools and districts. In addition, unless states adopt the same data definitions and coding systems, drawing valid comparisons among states will be difficult if not impossible, and will, at a minimum, require the time-consuming process of crosswalking the data elements from one state system to another state system.

District difference

Many districts have their own systems for coding data elements such as "course." Local education agencies may either adopt the state's system or School Codes for Exchange of Data (SCED), or map existing local codes to the state or SCED ones, an approach that may be less labor-intensive than changing standards.

For instance, if one state counts as "graduates" students who dropped out of high school, but later earned a GED or enrolled in a postsecondary institution, that state's cohort graduation rate may be higher relative to a state that considers such students "dropouts." Common codes for recording education data (academic courses, exits, attendance, etc.) or, alternatively, the crosswalking of local codes to a common system offers many benefits. The establishment of common data codes like those for race and ethnicity or for courses makes it easier to transfer information and draw comparisons across entities. In the case of course codes, for instance, a universal system can reduce time spent interpreting course information from transfer schools and help place students in the appropriate courses when they switch schools in the state or even across the country. For example, common course codes based on academic standards taught in each course will ensure that a student who completed algebra I in one district will be placed in the appropriate follow-up course in any school across the state. Use of common course codes also allows more reliable comparisons of performance data over time and across institutions. For example, an analysis considering the effects of taking algebra I will yield reliable results only if the courses measured are comparable in content—in other words, if they were classified by a common course-coding system based on academic standards taught in each course.

Maintain Metadata

Metadata, or "data about data," are critical to guiding proper data management and informed use. Without an organized approach to recording information about the agency's data elements, and without governing standards, the staff must remember or otherwise track all the information necessary to understand the data. This might include definitions; technical specifications such as field length and format, data source, due date(s), purpose, business rules, related calculations or transformations, and related policies; and all other information relevant to the creation, management, interpretation, and use of those data. In today's environment, where data elements are numerous, complex, and ever-changing, managing a data system without a robust metadata system would not be practical or advisable. In fact, many education agency staff consider a central, authoritative metadata repository a critical component of effective management and use of an agency's data.

Data models

A data model documents the agency's data architecture, helping users make sense of the many items that may be tracked by the system. By presenting an inventory of all individuals, places, and other entities involved in education, and by describing the relationships among them, an education data model can help education institutions, vendors, and researchers better understand the education data environment. Thus, a data model can help vendors and agency staff in data system design, or assist agency leaders select a product on the market that will meet their stakeholders' needs. A data model may also be a resource in the search for data to build into research designs.

While developing their data models, states and districts should focus on program area needs. IT should implement the model, but the business side of the organization should drive its design. Data models are offered by several vendors and by the federal government, which recently developed and continues to enhance a national data model called the National Education Data Model (NEDM). This nonproprietary model is available at *http://nces.ed.gov/forum/datamodel/index.aspx.*

National Standards Resources

State and local education agencies use a variety of data "standards," be they locally developed or adopted from state or national sources. Whatever the source, these standards are commonly set to meet federal and state data reporting requirements. However, shared data standards across the education community offer many other benefits. For instance, they enable interoperability, eliminating redundant data entry, lessening reporting burdens, reducing data errors, and facilitating data transfers. Using shared data standards also enables portability of student data and transcripts (see chapter 10 of *Book Two: Planning and Developing an LDS);* and common adherence to shared data standards allows valid data comparisons across district and state lines.

Education agencies may refer to several major sources for national data standards when designing, buying, adjusting, or using a data system. Figure 4 presents the three types of standards described above. Examples of what each type of standard looks like in reality are presented in the second column.

Common Education Data Standards Initiative

The Common Education Data Standards (CEDS) Initiative is a national, collaborative effort to identify a set of data elements of particular importance in K–12 education and the transition to postsecondary education. The CEDS Initiative has identified key data elements describing demographics, program participation, course information, and other attributes of students and the education system, as well as elements needed for high school-to-postsecondary transcripts and high school feedback reports. For each of these elements, the CEDS Initiative has defined detailed data standards including definitions, code sets, and a range of technical specifications. Most of these data standards have been drawn or adapted from established sources such as the NCES *Handbooks*, Schools Interoperability Framework (SIF) Association specifications, and Postsecondary Electronic

Figure 4. Data standards examples by type

Data definitions and code sets

Hispanic or Latino Ethnicity = An indication that the individual traces his or her origin or descent to Mexico, Puerto Rico, Cuba, Central and South America, and other Spanish cultures, regardless of race. The term "Spanish origin" can be used in addition to "Hispanic or Latino."

Code set:

02304 - Hispanic or Latino

02305 - Non-Hispanic/Latino

(Source: NCES Handbooks, Version 7.0)

Technical specifications

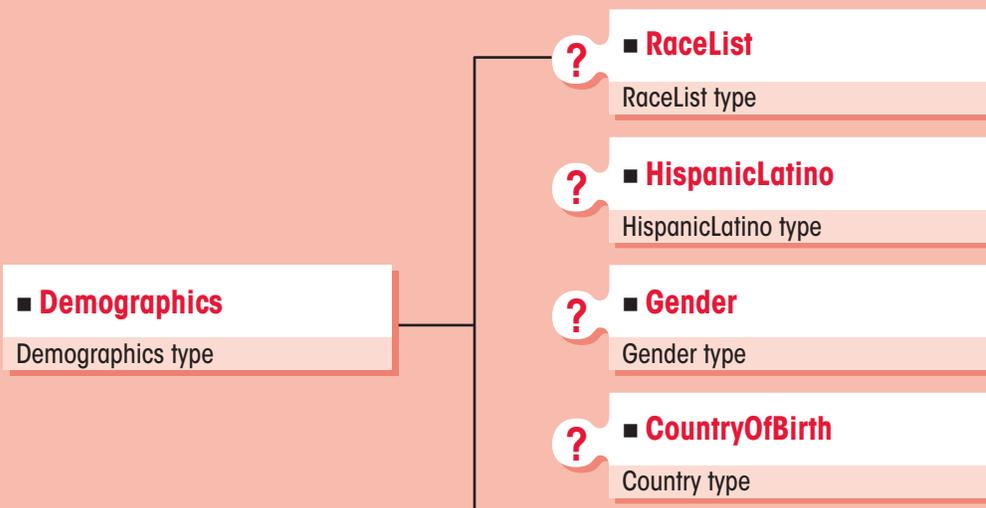| Element/@Attribute	Char	Description	Type
HispanicLatino		An indication that the individual traces his or her origin or descent to Mexico, Puerto Rico, Cuba, Central or South America, or other Spanish cultures, regardless of race.	Values: Yes No

Table 6.2.32-1: HispanicLatino

<HispanicLatino>Yes</HispanicLatino>
Example 6.2.32-1: HispanicLatino

(Source: SIF Specification, Version 2.3)

General guidelines and relationship

Demographics
Demographics type

? ■ **RaceList**
RaceList type

? ■ **HispanicLatino**
HispanicLatino type

? ■ **Gender**
Gender type

? ■ **CountryOfBirth**
Country type

(Source: SIF Specification, Version 2.3)

Standards Council (PESC) schema; and will be incorporated into the National Education Data Model (NEDM). Ideally, this core set of data elements will be voluntarily adopted by education agencies and marketplace providers. These sources are described on the following pages.

For more information on the CEDS Initiative, visit *http://commondatastandards.org.*

ED*Facts* initiative

ED*Facts* is a data initiative of the U.S. Department of Education compiling national, K–12 education data by consolidating previously separate federal collections. By combining these collections, ED*Facts* aims to centralize performance and other aggregate data for decision- and policymaking in order to streamline submissions to the federal government and eliminate redundancies, thus easing the burden on state agencies. Data collected for ED*Facts* include student and staff demographics, program participation, student performance and completion, school and district directory data, revenues and expenditures, school choice options, and other information. As a standards resource, ED*Facts* provides data elements, definitions, and code sets. The Department of Education also publishes technical specifications for ED*Facts* to guide the file submission process. Since much of the data collected by states are used to meet federal reporting requirements, the standards provided by ED*Facts* are commonly adopted by the states to facilitate compliance. In fact, all of its data elements have been incorporated into the National Center for Education Statistics (NCES) *Handbooks,* the Schools Interoperability Framework (SIF) specifications, and the National Education Data Model (see below). States submit their ED*Facts* data to the U.S. Department of Education through The Education Data Exchange Network (EDEN).

For more information on ED*Facts,* visit *http://www.ed.gov/about/inits/ed/edfacts/index.html.*

NCES *Handbooks*

The NCES *Handbooks* include a vast collection of basic data elements and option sets. The resource's stated purpose is "to provide a comprehensive listing of all data elements that might be needed for decisionmaking related to managing an education system, reporting to state and federal education agencies, and computing indicators of school effectiveness." Data elements are updated annually and organized into seven "domains" or levels: class, intermediate educational unit (IEU), local education agency (LEA), school, staff, state education agency (SEA), and student. For each data element, a definition is provided along with an option set if applicable. As a standards resource, the *Handbooks* offer a catalog of data elements, definitions, and code sets that is consistent with all the data elements needed to submit to the ED*Facts* data collection. They also include the School Codes for the Exchange of Data (SCED) and several standards provided by Forum publications (see below). Additionally, most of the *Handbooks'* terms, definitions, and code sets have been incorporated into the SIF specifications and the NEDM (see below).

For more information and to access the NCES *Handbooks Online,* visit *http://nces.ed.gov/programs/handbook.* Education agency administrators may also use the *Handbooks'* Data Dictionary Customization site to build their own data dictionaries.

The *Handbooks* include the Secondary School Course Classification System: School Codes for Exchange of Data (SCED), which presents a course taxonomy and course

descriptions for secondary education.* These codes are specifically intended to help education agencies track students longitudinally as they advance through grade levels, transfer to different schools, or enroll in a postsecondary institution. (See appendix C for additional information on the SCED.)

National Education Data Model

The National Education Data Model (NEDM) provides general data guidelines, depicting the relationships between a large collection of data elements collected and used in P–12 education. Specifically, NEDM focuses on the granular data items, attributes, and relationships associated with teaching, learning, and business operations at the school and district levels. For instance, the data model will tell you that a student with a specific name, physical address, phone number, displacement status, and other attributes, receives services from a teacher and participates in a class that has a room number within a building, which is a capital asset defined in the NCES Handbooks, and so on. Version 2 of the data model can be accessed at *http://nces.ed.gov/forum/datamodel/index.aspx.* The model currently includes all of the elements contained in the NCES *Handbooks,* and overlaps considerably with the data elements in the SIF specifications (see below). Several Forum standards are also included (see box below). In the future, NEDM is also expected to include elements from postsecondary education that are included in the Postsecondary Electronic Standards Council standards (see below).

National Forum on Education Statistics

The National Forum on Education Statistics (the Forum) has long been a leading resource for education data standards, focusing on issues of data standardization and basic data elements. A group of state and local education agencies, the federal government, and other organizations, the Forum has produced several guides that provide voluntary, best practice recommendations about data standards, including definitions, codes, and education data system components. To date, these products have covered areas such as crime, violence and discipline, attendance, exits, finance, facilities, and student displacement. Many of the Forum's standards have been incorporated into the NCES *Handbooks,* the NEDM, and the SIF specifications. For more information on the Forum's publications, see appendix F.

---

 *A similar collection of course codes for the elementary and middle school levels is currently being developed by the Forum. For more information, see the Forum's Prior-to-Secondary School Course Classification Working Group at *http://nces.ed.gov/forum/emscourseclass.asp.* For a basic list of elementary-level course codes in the NCES *Handbooks Online,* see the related options list for the NCES *Handbooks* data element "Elementary Course/Subject" at *http://www.nces.ed.gov/programs/handbook/elementinfo.asp?elementid=5491.*

Postsecondary Electronic Standards Council

The Postsecondary Electronic Standards Council (PESC) is an organization of colleges and universities; professional and commercial organizations; data, software, and service providers; nonprofit organizations and associations; and state and federal government agencies. Among the organization's missions is to create data standards to facilitate the exchange of data among postsecondary institutions. As a standards resource, PESC provides a range of standards for higher education, listing data elements, definitions, and code sets; and specifying technical requirements. The PESC standards for student transcripts have been crosswalked to the SIF specifications for student records to ensure comparability and completeness. However, because the standards are implemented differently, some variations exist and the two organizations continue to work together to ensure interoperability. PESC data elements related to student transitions to postsecondary education, such as e-transcript information, will also be included in the NEDM in the future. To enable data sharing with postsecondary institutions, K–12 education agencies may use PESC standards about students bound for, or enrolled in, higher education.

For more information on PESC, visit *http://www.pesc.org.*

## Schools Interoperability Framework Implementation Specifications

The Schools Interoperability Framework (SIF) Association is a nonprofit organization whose members include local and state K–12 agencies, software vendors, and others in the education community. The organization has created and continues to enhance a vendor-neutral "technical blueprint" for exchanging K–12 data. As a standards resource, SIF offers a full range of standards and defines suggested standards for naming, defining, and formatting data elements; as well as the technical specifications to allow software applications from different developers to easily interact and exchange data. SIF also includes a data model that depicts the relationships among the data; and data elements in areas such as student information, assessment, facilities, finances, food services, transportation, and professional development. The SIF specifications incorporate the NCES *Handbooks* elements and code sets whenever possible. All of the ED*Facts* elements are also included, and the SCED codes are referenced, as well as other standards provided in Forum products. SIF also overlaps with PESC standards related to transitions to postsecondary education, such as student transcript information.

For more information on the SIF specifications, visit *http://www.sifinfo.org/us/index.asp.*
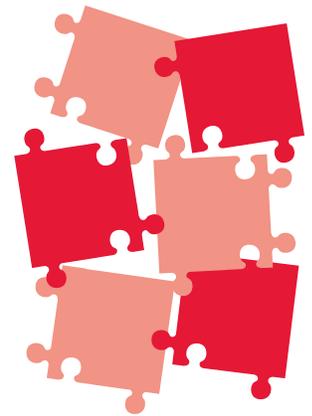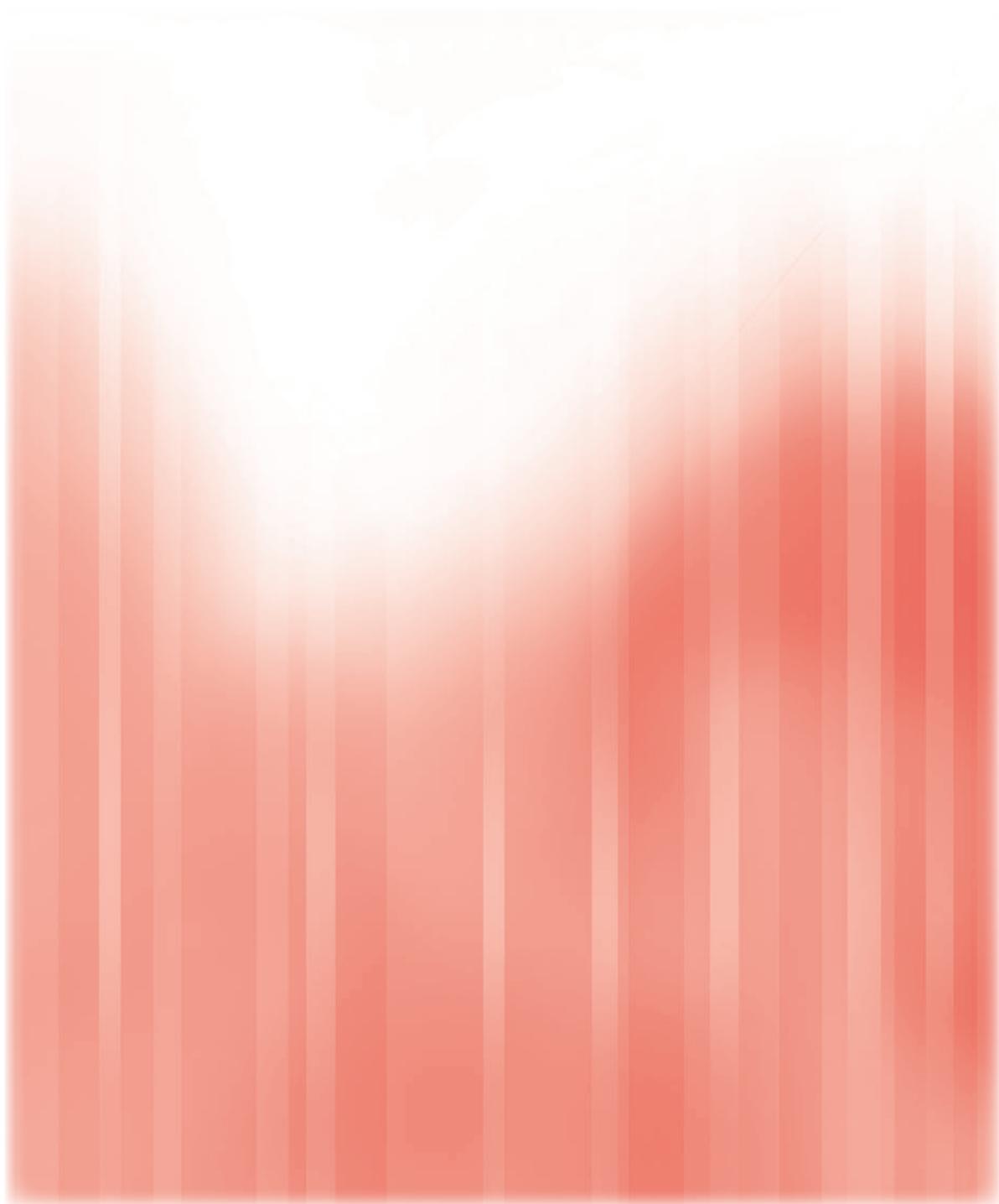
Securing the Data, Protecting the Individual
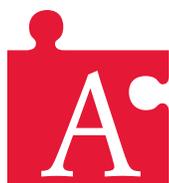
As state and local education agencies implement LDSs, the collection, management, and dissemination of individual student records will increase the need to protect individual privacy and dramatically raise the stakes for data security. While many state agencies, districts, and schools have not maintained student-level longitudinal data in the past, many have handled extensive and sensitive records on individual students and staff. And even the aggregate data sets maintained at the district and state levels may contain information that can be used to single out individual students, thus requiring special protection.

A number of federal and state laws and related regulations exist to protect the privacy of individuals, and these laws and regulations must be followed. Education agencies must also create and implement their own policies and procedures to guide staff in accordance with these laws and regulations, and these policies and procedures should be reinforced with staff training and technology solutions to protect sensitive records. The final two chapters address these issues.
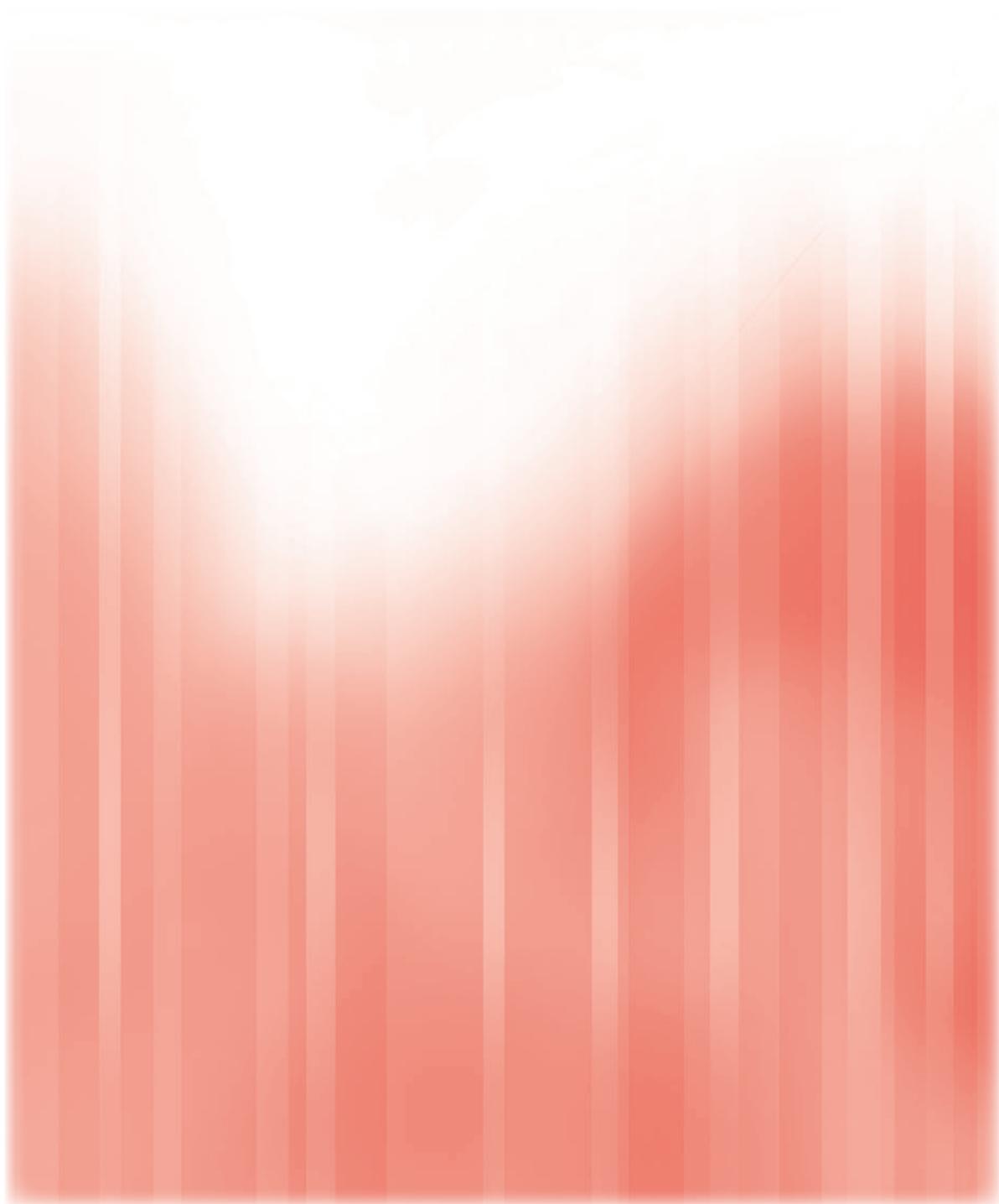
Chapter 7

PRIVACY AND CONFIDENTIALITY

Privacy ≠ Confidentiality

Though often confused, there is a distinction between privacy and confidentiality. "Privacy refers to an individual's right to withhold information, that is, not to divulge information to anyone else. Confidentiality refers to the handling of information that has been obtained by a second party."

(National Postsecondary Education Cooperative 1998)
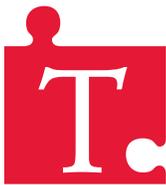
To reach their potential, LDSs must be used to collect, maintain, and make student- and staff-level data available to a wide variety of audiences. Teachers, students, principals, legislators, researchers, postsecondary administrators, and others can benefit from access to longitudinal data (see chapter 5 of *Book One: What is an LDS?)*. However, while these data can greatly enhance the ability to efficiently allocate resources and improve programs, instruction, and achievement, the sensitivity of personally identifiable information and the need to protect it cannot be overstated.

Individual privacy must be safeguarded in compliance with federal and state laws and regulations; and unauthorized and unlawful access must be prevented. Procedures should therefore be developed to allow secure and appropriate data sharing with organizations and users throughout the education community and beyond. While there has been debate and uncertainty over how best to protect privacy without limiting research and data access, many states have demonstrated that an effective balance can be struck. This chapter provides a basic overview of issues and relevant laws about data protection.

Federal Privacy Laws

Don't take it from us!

Information offered here on these federal laws should not be considered legally binding interpretations. Given the complex and dynamic nature of these laws, specific questions about student record confidentiality should be referred to the appropriate federal office (e.g., the Family Policy Compliance Office), or your agency's legal or administrative agents. For additional resources on privacy issues, see appendix C.

This chapter provides brief overviews of the four key federal laws that directly affect the data collected and maintained by education agencies. These are the Family Educational Rights and Privacy Act (FERPA), which applies to the vast majority of education data; and the Health Insurance Portability and Accountability Act (HIPAA), the Individuals with Disabilities Education Act (IDEA), and the National School Lunch Act (NSLA), which apply to education data in some cases. Though the details of these laws and their official interpretations do not spell out every detail, they do provide basic guidelines on what data can be shared, with whom, and under what circumstances. State policies and laws often work out some of the implementation issues, and sometimes add further privacy protections.

Personally identifiable information

Before reviewing the privacy laws of import, which data are affected should be clarified. These privacy laws put no restrictions on data sharing if all individually identifiable information is removed from the records. According to the FERPA regulations, "personally identifiable information" includes, but is not limited to, the following:

    a. the student's name;

    b. the name of the student's parent or other family members;

    c. the address of the student or student's family;

    d. a personal identifier, such as the student's Social Security number, student number, or biometric record;

    e. other indirect identifiers, such as the student's date of birth, place of birth, and mother's maiden name;

    f. other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty (see below); or

    g. information requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates.

Source: *Federal Register* (Dec. 9, 2008)

In some cases, even when the more obvious personal information is removed from individual student records, users may still be able to match individuals to their records when those students' characteristics are rare or unique. For example, the only female Asian 3rd grader enrolled in a school will be easily identifiable in a data set, even if all of the obvious personally identifiable information is removed from her record. In cases like these, state and local staff need to take a proactive approach to preventing such invasions of privacy, using techniques such as perturbation, encryption, redaction, or deletion of data to maintain the confidentiality of private information.

NCES has more detailed information...

...about protecting education data:

- *Basic Concepts and Definitions for Privacy and Confidentiality in Student Education Records (2010)*
  http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011601

- *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting (2010)*
  http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011603

- *Data Stewardship: Managing Personally Identifiable Information in Student Education Records (2010)*
  http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011602

Agencies may also need to manipulate aggregate data sets or performance reports that include groups of less than a specified number of students–5 or 10, for example–to avoid exposing an individual student's score or other personal information. In practice, agencies may choose to suppress all the information about a small subgroup, or combine subgroups to raise the number or percentage of students reported in a group. This minimum n should be large enough to protect privacy and ensure statistical reliability, while also avoiding the loss of too much detail (ESP Solutions 2008). Similarly, agencies should also manipulate their data sets or reports if certain percentages are too large. For instance, if 100 percent of students in a school are eligible for free or reduced-price meals, users will know the eligibility status for every student in that school. In this case, the percentage may be artificially decreased to create uncertainty about who is eligible and, thus, protect students' privacy.

Anonymized data are previously identifiable data that have been de-identified and for which a code or other link no longer exists. An investigator would not be able to link anonymized information back to a specific individual.

(http://healthcare. partners.org/phsirb/ hipaaglos.htm)

Once personally identifiable information has been removed or manipulated as necessary, the resulting anonymized data may be shared with the public without consent according to FERPA. However, some states restrict access even to these anonymized data to varying degrees.

Family Educational Rights and Privacy Act

The Family Educational Rights and Privacy Act of 1974, commonly referred to as FERPA, is a federal law intended to protect the privacy of student education records. The law applies to all education institutions that receive federal funding under programs administered by the U.S. Department of Education.

FERPA has increasingly become an important issue in the education community, especially recently because of the emerging implications of LDS development and data sharing. As a result of the rapid advance of technology and the expansion in data collection and demand, a rising level of uncertainty has surrounded the law's implementation. FERPA generally prohibits agencies from sharing personally identifiable information without written consent (though a number of exceptions are made), and many agencies have been reluctant to share data in some instances for fear of infringing on their students' rights. While this hesitancy is often justifiable, in some cases agencies may be overly cautious and withhold

information based on too strict an interpretation of the law. This roadblock to data access has been a continuing source of frustration for many potential users, primarily education researchers. And, it might be possible to use FERPA as an excuse not to release data that might portray the education system in an unfavorable light (Viadero 2006).

Written when most individual education records were maintained on paper at the local level, FERPA's authors did not consider modern electronic records or statewide LDSs. To keep up with the evolution of technology and culture, the U.S. Department of Education has offered subsequent interpretations of FERPA, allowing the education community to progress while still honoring the law. In 2008, for example, a revised interpretation of FERPA was issued to clarify many of the ambiguities and remove some of the roadblocks that existed in previous regulations. Of major significance were expanded disclosure rights to state education agencies, effectively paving the way for easier access to statewide student-level data (previously, only districts were granted disclosure rights, a limitation that, among other concerns, hindered researchers seeking to compile significant samples of student data). Additionally, the new regulations refined guidance concerning disclosure of student information to parents, third parties, former schools, state auditors, and research institutions; recordation (recordkeeping for each disclosure); data sharing among K–16 education institutions; de-identification of shared records; and the use of Social Security numbers.

While the new regulations were intended, at least in part, to strike a balance between the protection of student privacy and the facilitation of valuable research, questions about the law remain. These uncertainties center primarily on the particulars of sharing P–12 data with researchers, postsecondary institutions, students' former schools or districts, and other state agencies such as workforce and social service agencies (Education Counsel 2008). Further clarifications may be necessary to reconcile the law with the federal government's goal of fostering the development and effective use of statewide, student-level LDSs. For more information on FERPA, visit *http://www2.ed.gov/policy/gen/guid/fpco/index.html*.

Health insurance portability and accountability act

The Privacy Rule of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) is intended to protect the confidentiality of individual health records. In general, elementary and secondary schools and districts are not subject to HIPAA, because even if they qualify as a "covered entity," any health-related data they maintain are considered "education records" subject to FERPA. FERPA takes precedence even for records created by school nurses or other healthcare providers, if they are under the direct control of the school. Most schools and districts must comply with HIPAA only when they request medical records from an outside health care provider. Once those data are received from the outside health care provider and in the education institution's possession, they are considered education records and become subject to FERPA. Private schools that do not receive funding from the U.S. Department of Education are the most common exception. In this case, any health-related data about students or others who receive health care services are considered "protected health information" and must be protected in compliance with HIPAA. For more information on HIPAA, visit *http://www.hhs.gov/ocr/privacy/index.html*.

Individuals with Disabilities Education Act

Records on students in special education programs are subject to the privacy requirements of the Individuals with Disabilities Education Act (IDEA). The IDEA requirements include many of the same protections provided by FERPA, with a few differences related to the handling of student records and several additional requirements. For instance, information
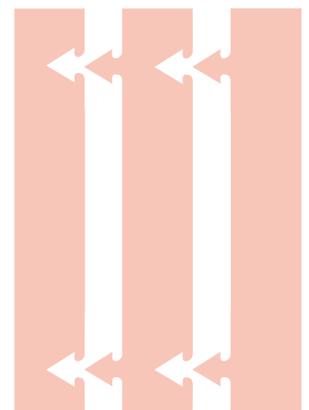
on a student's disability cannot be shared without parental consent. Institutions subject to both FERPA and IDEA must comply with the privacy provisions of both laws. Considerable overlap between the laws simplifies this task. For more information on IDEA, visit *http://www2.ed.gov/about/offices/list/osers/osep/policy.html.*

National School Lunch Act

Data on students' eligibility for free and reduced-price meals, and information obtained as part of the National School Lunch Program of the U.S. Department of Agriculture, are covered by confidentiality restrictions in the National School Lunch Act (NSLA). While also subject to FERPA, the privacy restrictions of the NSLA are stricter in two cases: free and reduced-price meal eligibility. The sharing of individually identifiable information obtained during the eligibility process is, with some exceptions, prohibited without parental consent. However, in some cases, eligibility and other information about the student's household may be shared with select individuals and programs, such as some assessment programs (e.g., the National Assessment of Educational Progress). In most states, though, these data may be made available to users if all personally identifiable information has been removed. For more information on NSLA, visit *http://www.fns.usda.gov/cnd/lunch.*

State Laws

Many states have established their own laws and policies that either mirror, or expand on, the basic guidelines provided by federal laws. For instance, some states have issued laws dealing with areas within FERPA they considered ambiguous. They may, for example, have defined authorized disclosures more specifically, established a process for approving disclosures through written agreements, specified roles and responsibilities for protecting privacy, or allowed the use of Social Security numbers as student identifiers. Other states have passed laws that explicitly permit certain data sharing between the K–12 and postsecondary sectors, among state education agencies, or with other state agencies such as workforce or social service agencies. On the other hand, some states have enacted laws that are more stringent than the federal laws. For instance, they may prohibit data sharing that would be permitted under the current interpretation of FERPA, such as disclosures from the state education agency to districts receiving a transfer student, or to teachers about their students. To ensure a balance is struck, states should review their existing privacy laws, regulations, and guidelines so that they will not inhibit effective use of the student-level data they intend to make available through their LDS.

(Sources: DQC 2007, Hill 2008, and Nunn et al. 2006)

DATA SECURITY
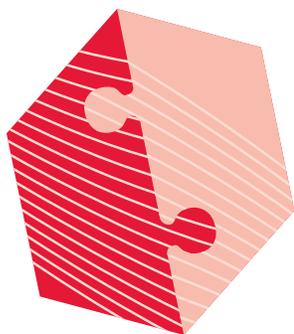
W hether or not an education agency has an LDS, data need to be secured to prevent unauthorized access and tampering. However, the collection, maintenance, and dissemination of student-level data via an LDS and other source systems increases the importance of data security. While many districts have long stored some personally identifiable information, states take on a new responsibility when they begin to manage personally identifiable, student-level data.

The Forum has more detailed information…

…about data security issues:

- *Forum Unified Education Technology Suite (2005)*
  http://nces.ed.gov/forum/pub_tech_suite.asp

- *Forum Guide to Decision Support Systems: A Resource for Educators (2006)*
  http://nces.ed.gov/forum/pub_2006807.asp

Security measures should keep sensitive data out of the wrong hands, while allowing maximum accessibility to users. An LDS contains sensitive data that can be compromised and used to expose restricted personal information, thereby violating privacy (Houde 2008). Protections must allow access to authorized users, while barring others from seeing or manipulating the data.

Know your data

Ownership model of data security

Within a data governance structure, ultimate responsibility for each of the agency's data elements should be assigned to a single data steward (see chapters 1–3). These stewards should work with the security team to determine the sensitivity of, and appropriate level of security for, every data element. Together, they may

- classify each item's level of confidentiality (e.g., public or restricted); and
- identify the user groups, by characteristics such as job function or "need to know," that should be granted access to each element.

As discussed in chapter 3, data stewards should also be the point of contact for requests of data they manage, and may authorize data sharing in response to those requests.

The first step in securing an education agency's data is developing a clear view of its information "landscape."

➤ Determine what the agency has. Take an inventory of all of the data the agency collects and maintains.

➤ Locate all the data. Document where the data are stored, including servers, individual computers, filing cabinets, and other media such as CDs and storage devices.

➤ Document the "ownership" of each data element. Each data element should be the responsibility of a single data steward, who should be clearly identified.

➤ Determine the sensitivity of each data element the agency manages based on privacy requirements under state and federal laws. Document the risks associated with exposure of sensitive information and/or assign a risk level to each element, perhaps through a simple rating scale.

➤ Document who has access to what data, including internal staff, contractors, vendors, and external users. Note who can do what with the data (e.g., manipulate vs. view) and record all the ways these authorized users can access the data.

➤ Determine and document how long various users will be allowed to access data and ensure access is denied once it is no longer appropriate. For example, agencies

must diligently end or adjust staff access to data when they change positions or leave the agency. Also, states or districts (or school policymakers) must determine if teachers will be allowed to access personally identifiable information on their former students (some agencies update access by enrollment annually, while others grant teachers broad access to their former students' information for a year or more, even into postsecondary).

➤ Document data sources. From what source systems do your data originate? Who sends them?

➤ Document data recipients. These may include federal agency departments, postsecondary institutions, research organizations, and all others who receive the data.

➤ Document how data are transmitted. Data might be transmitted electronically, mailed on paper, etc.

Role-based data access

Users should have easy access to the data in an LDS, and standardized reports showing aggregate data and analysis results should be publicly available. Additionally, for ad-hoc querying and analysis, the general public may be given access to aggregate statistics and to non-identifiable individual student records. For personally identifiable information, users should be granted varying levels of access depending on their role, needs, and responsibilities. For instance, through an online application, users could gain access to permitted information by signing on with their individual username and password. A student's record may be made available to that student, as well as to his or her parents or guardians, current teachers, counselors, school, district administrators, etc. However, the specific information shared may vary depending on the user's identity; for example, only parents might be allowed to see a student's lunchroom account balance. Researchers with appropriate contracts and permission may also be granted access to some personally identifiable data.

The Data Policy Committee or Data Governance Committee should document accessibility by user roles, as well as for what purposes the data can be used. These specifications should be granular enough to detail the rights of each type of user, allowing access to the information each is entitled to, but no more. Also, while access rights may be determined at the state level, some software programs allow delegation of access rights to occur at the local level. In other words, the application allows "delegated" district administrators to grant certain users access to sensitive information, if they deem such access appropriate.

While some education agencies run multiple software systems to access separate databases or data sets, others implement "single sign-on" systems to manage user identity and streamline access. Rather than assigning each user a different password and username for each data system, these agencies maintain a single username and password for each user and grant appropriate access across applications. Each user's access rights can be tailored based on his or her specific need for the data in each system.

Keep Only the Data Your Stakeholders Need

The more data an agency has, the more it must secure. Although stakeholders may demand a wide range of information, each agency should consider disposing or securely archiving any data deemed unnecessary, especially if they contain personally identifiable information. In this stage of the life cycle of information (see chapter 1 of *Book Two: Planning and Developing an LDS),* data should be destroyed in a manner consistent with their sensitivity.

➤ Completely remove information ("wipe") from old computers and storage devices before disposing of them; and shred, burn, or pulverize unnecessary paper records. Remote and onsite staff should follow the same procedures.

➤ Going forward, collect only the information required to meet business and stakeholder needs.

➤ Create a record-retention policy that details what information should be stored by the agency, how it should be secured, how long it should be kept, and how it should be destroyed once it is no longer needed. Agencies may set up a formal review process to assess data's value and authorize disposal.

Secure the Data
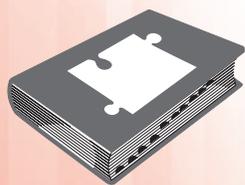
Authentication is the verification of a user's identity through means such as the submission of a unique password and/or other personal information. **Authorization** is the mechanism by which that authenticated user is granted access rights (e.g., the right to view data of varying degrees of sensitivity, or the right to manipulate data in addition to viewing them).

Agencies need to stay on their guard, identifying vulnerabilities and adapting to ever-changing security threats. Threats may come from within the organization or outside agency walls. The Internet, for instance, increases the risk to student privacy as individuals from the local education community or from across the globe can hack into data systems to change test scores, unleash viruses, or just wreak general havoc.

➤ Establish a group or office specifically focused on security issues, perhaps creating an enterprise-wide security plan; and implement security strategies to manage data access and use.

➤ Assign a security officer to lead this office. This staff member should be well-versed in all relevant privacy laws, and with the technology and business processes that facilitate compliance. The security officer may coordinate the agency's security plan (authentication, intrusion detection, etc.) and must ensure that all staff are appropriately trained to protect the agency's data.

➤ Store data in a secure location accessible only to authorized personnel, and lock when not in use. This would include all sensitive information contained on servers, computers, media such as CDs, or on paper.

➤ Automatically encrypt hard drives and use only password-protected thumb drives

for transferring sensitive data (Houde et al. 2007).

➤ Set access controls to the network and review them periodically. A user's identity should be authenticated with a complex password, pass phrase, or other personal information.

➤ Based on the level of access determined by staff (e.g. data stewards) and implemented through technology, each particular user should only be given access to authorized information.

➤ Use intrusion-detection systems to identify suspicious access to the network.

➤ Establish and utilize infrastructure components such as firewalls, backups, and antivirus and antispyware software (EIMAC 2008).

➤ Protect data while they are moving ("in motion") between data systems or to data users. For instance, student-level data might be encrypted before they are fed from a source system into the LDS, or from the state agency's LDS back to a district.

➤ To help keep private data from getting into the wrong hands, convene a Data Request Review Board (see chapter 3) to establish a clear process for handling data requests in an orderly and consistent fashion.

➤ Create a contingency plan to facilitate a quick and appropriate response in the event data security is threatened or breached. This plan should specifically describe responses to a range of scenarios. For instance, how will the agency respond to network intrusion, a stolen laptop, or wrongful dissemination of sensitive information? Whom will you notify–law enforcement, staff, the individuals whose personal data have been compromised, the public, etc.–and should you use the phone, email, or other means of communication? How will damage to the system be controlled? How will the impact of the breach be assessed?
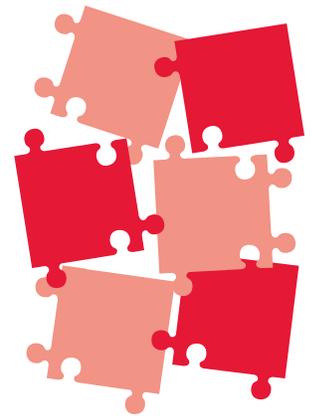
Even with a solid security plan, agency data will not be secure without proper implementation. All staff, not just IT, should understand the sensitivity of the records and the vulnerabilities of the system, and security should be a priority in everyone's daily routine. Improving agency security therefore involves a certain degree of culture change.

Disaster preparedness

Natural and manmade disasters can severely disrupt educational activities, displacing students and interrupting services including data collection, maintenance, and use of data. LDSs, which may consolidate a wide range of data and data processes, are likely to be necessary for carrying out day-to-day business ("mission critical"). Moreover, when disaster strikes, these systems are vital to mitigating the effects of the crisis. For instance, the LDS can be used for enrolling displaced students in the appropriate grades, courses, and programs; meeting accountability requirements; and efficiently allocating funding. Agencies should carefully plan for destructive or disruptive events, including physically safeguarding the system, designing the system architecture to facilitate displaced student tracking (e.g., including data elements such as displacement identifiers and event descriptors), and creating policies for tracking students and exchanging data after a crisis.

The implementation of a state-level LDS offers several clear advantages in a crisis. For example, consolidating agency data into a single data store will facilitate preparations for a disaster, as the need to modify and coordinate a multitude of silo systems would be limited. During and after a crisis, the efficient exchange of high-quality, student-level data that are verified by centralized procedures and marked by individual identifiers via an LDS may help in many ways. For instance, administrators can use the data to efficiently and precisely target resources, provide displaced students with the proper services, and pass the scrutiny of federal data audits. Aggregate counts, on the other hand, may be inaccurate as they often contain duplicate counts. They may also be untimely as they take more time to produce, leading to slow and inefficient distribution of resources and a haphazard provision of services. Finally, states with an LDS may be able to ease the reporting burden of districts needing federal aid, as they may be able to deal directly with federal agencies on school districts' behalf.

(National Forum on Education Statistics 2010)

Additional resources:

Disaster preparedness

■ *Crisis Data Management: A Forum Guide to Collecting and Managing Data About Displaced Students* (2010)

http://nces.ed.gov/forum/pub_2010804.asp

This guide focuses on data issues surrounding the displacement of students. It offers best practices and lessons learned regarding planning activities that can help mitigate the disruptive effects of crises on education data systems and business continuity.

■ *Disaster Prevention and Recovery for School System Technology* (ESP Solutions 2005)

Available at *http://www.espsolutionsgroup.com/resources.php.*

This white paper offers best practices related to disaster recovery planning.

■ *IT Disaster Recovery and Business Continuity Toolkit: Planning for the Next Disaster* (NASCIO 2007)

http://www.nascio.org/publications/documents/nascio-drtoolkit.pdf

This toolkit provides a framework to assist CIOs and other agency leaders in the development of IT disaster preparedness plans. It includes a collection of high-level checklists and questionnaires to help focus preparation efforts.

Train and Inform Data Handlers

The Forum has more detailed information…

…about data ethics:
- *The Forum Code to Data Ethics (2010)*
  http://nces.ed.gov/forum/pub_2010801.asp

- *The Forum Guide to Data Ethics Online Course (2010)*
  http://nces.ed.gov/forum/dataethics_course.asp

With an LDS, more staff members will gain access to sensitive student-level data. As access to data expands, so must security and confidentiality training to avoid unlawful data sharing or use. Some best practices follow.

➤ Establish a training plan that tailors instruction to staff with different levels of access to sensitive data, giving those with more access more rigorous training. Training may also be tailored to specific groups of data users based on job functions.

➤ Monitor the access granted to staff and provide additional training as needed.

➤ Train and retrain staff and contractors periodically on the proper and ethical handling of sensitive data. This training should be completed before access to sensitive data is granted.

➤ Hold staff accountable for failures to adhere to the agency's security procedures and confidentiality policies.

➤ Require employees and contractors to sign nondisclosure agreements.

➤ Require researchers to sign memoranda of understanding (MOU) that detail privacy and security requirements.

➤ Establish a means for communicating security issues to staff, perhaps via a security website, email newsletters, or meeting updates from security staff; communications should be both periodic and impromptu as issues surface.

➤ Specify security requirements in requests for proposals (see chapter 14 of *Book Two: Planning and Developing an LDS)* and assess prospective vendors' security practices or software specifications; this will help ensure the service provider or product can, in fact, meet the agency's security needs.

LDS Lore:
Identity theft in the printer room

At the school district office, Margaret typed in her password and accessed the teacher information system. She found the data she needed and sent it to the printer. The phone rang—Sally was calling about lunch. Starving, Margaret grabbed her coat and headed out for a burrito. Meanwhile, at the printer, Bonnie lifted a stack of unclaimed paper from the tray and set it on the table. She waited for her job to collate, snatched it, and left the room. After lunch, Margaret was extremely busy as one meeting flowed into another until it was time to make the commute home. Before close of business, Chris, the district's data governance coordinator, was making his rounds and noticed the pile of papers on the table. He picked them up, quickly realized what they were, and raced back to his office.

Meanwhile, Richard, the new janitor, was buttoning up his uniform. He'd heard that staff at the district office sometimes leave pretty valuable information lying around (the kind he'd read about in that non-disclosure agreement he'd signed) and he was looking forward to getting his hands on some. Later, at the office, he steered the vacuum around the desks, down the empty halls, and into the printer room, keeping his eyes peeled. He checked the recycling bins, garbage cans, table tops, looking everywhere for names, Social Security numbers, dates of birth—the good stuff that he could turn into cash. But even though he cleaned the place from wall to wall, nothing turned up. Better luck tomorrow night, he told himself before turning off the lights.

The following morning, Sally opened her email to find a message from Chris calling a mandatory meeting for all data stewards. The subject: Safeguarding sensitive data. A second email came from the district's security officer announcing a new round of training on data security. Just then, it all came back to Sally: the print job...the Mexican food...the meetings. Sally thought, as she rushed to the meeting, "I hope I didn't cause any damage."

Appendix A
Overview of the LDS Guide Series

The *Traveling Through Time: The Forum Guide to Longitudinal Data Systems* series consists of four guides intended to help state and local education agencies meet the many challenges involved in developing a robust longitudinal data system (LDS), populating them with quality data, and using this new information to improve the education system.

   Book One of Four: What is an LDS? focuses on the fundamental questions of what an LDS is (and what it is not), what steps should be taken to achieve a sound and successful system, what components make up an ideal system, and why such a system is of value to education. *Book Two of Four: Planning and Developing an LDS,* discusses the critical planning and development phases of an LDS project, from stakeholder engagement and needs assessment all the way through to system evaluation. *Book Three of Four: Effectively Managing LDS Data* explores several fundamental challenges of data management, focusing on data governance, data quality, privacy, and security. Finally, *Book Four of Four: Advanced LDS Usage,* will address the effective utilization of LDS data, discussing the users and uses of the data; and emphasize the need for effective training and professional development. The figure on the next page lays out the major issues covered in each of the four books in this Forum guide series.

Book I: What Is an LDS?

- Understanding what an LDS is (and is not)
- Appreciating the organizational steps needed to institute and effectively use an LDS
- Identifying the technical features and capabilities of an effective LDS and the additional features that can enhance the system's utility
- Recognizing the benefits of an LDS

Book II: Planning and Developing an LDS

- Engaging stakeholders
- Describing the current system
- Envisioning the desired system
- Defining needs, including data and functionality
- Gaining buy-in and funding
- Building relationships
- Writing an RFP
- Building or buying a system or components
- Transferring knowledge (e.g., from developers to staff)
- Defining and measuring success
- Refining the system

Book III: Effectively Managing LDS Data

- Defining governance structure
- Defining roles and responsibilities
- Collaborating to improve data quality and streamline operations
- Managing changes to the system
- Training staff to ensure data quality
- Auditing/validating data at all levels
- Establishing/following data standards
- Securing data to protect privacy
- Providing users access to key data

Book IV: Advanced LDS Usage

- Collecting, storing, and delivering key data
- Developing useful reports to fulfill common data requests and needs
- Developing user-friendly data tools to facilitate access and analysis
- Training users to utilize the technology
- Building awareness, understanding, and analytical capacity

Book One: What is an LDS? is available online at: *http://nces.ed.gov/forum/pub_2010805.asp*

Book Two: Planning and Developing an LDS is available online at: *http://nces.ed.gov/forum/pub_2011804.asp*

To order a print copy, please contact:
ED Pubs
U.S. Department of Education
P.O. Box 22207
Alexandria, VA 22304

Or call toll free 1–877–4ED–PUBS or order online at *http://www.edpubs.gov.*

Appendix B

References

Aarons, D. (2009). Leading the Charge for Real-Time Data. *Education Week*, June 3, 2009. Available at *http://www.edweek.org/ew/articles/2009/06/03/33dataleader_ep.html?tkn=PSN FXRv4MZsj0rDA8PJJQE7WcHpl%2F4ehVf78*.

Chatis Consulting. *Data Management Committee: Critical Data Issues*. (Unpublished.)

Data Quality Campaign (DQC 2007). *Maximizing the Power of Education Data While Ensuring Compliance with Federal Student Privacy Laws: A Guide for Policymakers*. Washington, DC: Data Quality Campaign (DQC).

Data Quality Campaign (DQC 2008). *Data Governance: Changing Culture, Breaking Down Silos, and Deciding Who Is in Control*. Washington, DC: Data Quality Campaign (DQC).

Data Quality Campaign (DQC 2009). *The Next Step: Using Longitudinal Data Systems To Improve Student Success*. Washington, DC: Data Quality Campaign (DQC).

Education Counsel (2008). *Needed Changes in FERPA Law, Regulations, or Guidance to Harmonize FERPA with the Functions of State Longitudinal Data Systems*. (Draft.) Available at *http://www.dataqualitycampaign.org/files/publication-needed_changes_ferpa_not_addressed_ regulations-122808.pdf*.

Education Information Management Advisory Consortium (EIMAC 2008). *The Need for Data Governance*. Retrieved from *http://www.ccsso.org/content/pdfs/EIMAC%20Brief%20 4%200308.pdf* on September 24, 2008.

ESP Solutions (2008). *Confidentiality and Reliability Rules for Reporting Education Data: A Guide for Establishing Decision Rules for Disaggregating and Reporting Assessment Results, and Other Indicators*. Austin, Texas: ESP Solutions Group.

Federal Trade Commission. *Protecting Personal Information: A Guide for Business*. Retrieved from *http://www.ftc.gov/bcp/edu/multimedia/interactive/infosecurity/index.html* on January 11, 2010.

Hill, E. (2008). *Redefining Student Data Access Policy*. Sacramento, California: Legislative Analyst's Office.

Houde, D.; Vocca, M.; Williams, D.; and Canada, B. (2007). *Security, Data Access*. Presentation given at November 2007 IES SLDS Grantee Meeting in Arlington, Virginia. Available at *http://nces.ed.gov/programs/slds/nov07_presentations.asp#sessionvi*.

Houde, D. (2008). *Arizona Department of Education Information Security Update*. Presented at the annual summer meeting of the National Forum on Education Statistics in Bethesda, Maryland.

Kansas State Department of Education (KSDE 2007). *Data Governance: The Kansas Approach.* Presentation given at the Education Information Management Advisory Consortium (EIMAC) Spring Meeting in Washington, DC.

Kansas State Department of Education (KSDE 2008). *Data Governance Program Version 2.2.* (Unpublished.)

National Forum on Education Statistics (2010). *Crisis Data Management: A Forum Guide to Collecting and Managing Data About Displaced Students* (NFES 2010–804). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

National Postsecondary Education Cooperative (1998). *Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education Data Collection, Retention, and Release* (NCES 98–280). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

NASCIO (2008). *Data Governance—Managing Information as an Enterprise Asset: Part I—An Introduction.* Lexington, Kentucky: NASCIO.

Nunn, J. and Harper Mainzer, K. (2006). *Longitudinal Data Systems: Summary of Current Issues and Potential Uses.* Retrieved from *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx* on December 4, 2008.

Redman, T. (2001). *Data Quality: The Field Guide.* Boston, Massachusetts: Digital Press.

Rozelle, R. (2006). *The Tennessee Statewide Longitudinal education Data System (SLeDS): The Governance Process.* Presentation given at the 2006 IES SLDS Grantee Meeting in Washington, DC.

Rozelle, R. (2008). *Role Descriptions and Responsibilities in the Realm of Data Governance.* Center for Educational Leadership and Technology. Available at *http://www.dataqualitycampaign.org/files/publication-role_descriptions_and_responsibilities_ in_the_realm_of_data_governance-022608.pdf.*

Schutte, S.; Edwards, S.; Fadaoff, S.; Ogle, T.; and Ligon, G. (2009). *The Process of Data Quality: State Discussion with Alaska, Wyoming, Missouri, and California.* Presentation given at the 22nd Annual MIS Conference in Seattle, Washington. As cited in The Process for Ensuring Data Quality, ESP Solutions (2009).

Tennessee Department of Education (draft 2008). *Data Manager Manual, Version 1.0.* Retrieved from *http://www.dataqualitycampaign.org/files/publication-data_manager_manual. pdf* on December 12, 2008.

U.S. Department of Education (2008). *F*amily Educational Rights and Privacy; Final Rule. *Federal Register, 73* (237). Available at *http://www2.ed.gov/legislation/FedRegister/ finrule/2008-4/ 120908a.pdf.*

Viadero, D. (2006). Scholars Cite Privacy Laws as Obstacle. *Education Week,* January 18, 2006.

Zhao, X.; O'Connor, B.; and Barroso, G. (2006). Objective of the Data Security Model. *InfoManagement Direct,* August 2006. Available at *http://www.information-management.com/infodirect/20060825/1061606-1.html.*

Appendix C
Additional Resources

Chapter 1. Data Governance

Data Governance: Changing Culture, Breaking Down Silos, and Deciding Who is in Control

Data Quality Campaign (DQC 2008).

http://dataqualitycampaign.org/files/meetings-dqc_quarterly_issue_brief-072908.pdf

This issue brief discusses the benefits of data governance, conditions helpful in successfully establishing data governance, and more. It also includes case studies of three states' experiences.

The Need for Data Governance

Education Information Management Advisory Consortium (EIMAC 2008).

Retrieved from *http://www.ccsso.org/content/pdfs/EIMAC%20Brief%204%200308.pdf* on September 24, 2008.

Data governance is driven by the need for data quality. This three-page brief discusses the components of good data governance, including leadership, data quality management, controlled analysis and reporting, security and confidentiality, resource management, and data use and accessibility.

The Next Step: Using Longitudinal Data Systems to Improve Student Success

Data Quality Campaign (DQC 2009).

http://dataqualitycampaign.org/files/nextstep.pdf

See page 10 for a section on data governance.

DataFlux Knowledge Center

http://www.dataflux.com/getdoc/3a87235c-d2b3-4c24-a794-6198669754da/data-governance.aspx

This page contains a host of white papers and other resources related to data governance. See the Data Governance Maturity Model for a good summary of the issue.

NASCIO Enterprise Architecture and Governance Committee

http://www.nascio.org/committees/ea

The Enterprise Architecture program of the National Association of State Chief Information Officers (NASCIO) was developed to enable the mission of state and local government. This web page contains a number of helpful resources on data governance, including a three-part series of papers on the topic.

The Data Governance Institute
http://www.datagovernance.com

This website contains an extensive collection of resources on data governance, including basic information, case studies, frameworks, books, and more.

The Data Administration Newsletter
http://www.tdan.com

This newsletter publishes many useful articles on data governance.

Management of an Education Information System
ESP Solutions (2005).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_EDInfo_System_Mgmt_ORG.pdf

This document describes how education agencies can organize resources for effective management of the information they acquire, store, process, and report.

Harnessing the Potential for Research of Existing Student Records Databases: An Action Agenda
National Center for Higher Education Management Systems (2005).

http://dataqualitycampaign.org/resources/31

This brief resulted from a meeting of academic researchers responsible for several state student unit record systems. Pages 6–13 discuss issues related to data governance.

Chapter 2: Basic Steps to Establishing Data Governance

Implementing Data Governance as the Foundation of a LDS Presentation
Tennessee Department of Education (2008).

http://dataqualitycampaign.org/resources/96

This presentation summarizes Tennessee's approach to implementing data governance, including roles, helpful strategies, and lessons learned.

Role Descriptions and Responsibilities in the Realm of Data Governance: Managing Information Assets
Tennessee Department of Education (2008).

http://dataqualitycampaign.org/resources/97

This document describes eight data management jobs and roles. It also includes a matrix of data governance activities, and the people/groups that need to be involved in each activity.

LDS Share
The following documents are available at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx:*

Data Management Committee Critical Data Issues Log
Tennessee Department of Education (2007).

This is a template for tracking critical data issues, which may be tracked by a Data Governance Committee.

Policies and Procedures Review: Data Management

Tennessee Department of Education (2006).

This document from Tennessee establishes policy and procedures governing data management and quality for the Department of Education. Roles and responsibilities of the data governance structure are also outlined.

Chapter 3: Groups, Roles, and Responsibilities of Data Governance

Role Descriptions and Responsibilities in the Realm of Data Governance: Managing Information Assets

Tennessee Department of Education (2008).

http://dataqualitycampaign.org/resources/97

This document describes eight data management jobs and roles. Also included is a matrix of data governance activities and the people/groups that need to be involved in each activity.

Data Governance Program, Version 2.2

Kansas State Department of Education (2008).

http://www.dataqualitycampaign.org/files/KSDE_Data_Governance_Program_Ver_2.2.pdf

This document details Kansas' approach to data governance. It includes an overview of the issue, and handbooks for the two major committees involved and for data stewards.

The DGI Data Governance Framework

Data Governance Institute.

http://www.datagovernance.com/fw_the_DGI_data_governance_framework.html

This resource presents a framework to help organizations classify, organize, and communicate complex activities related to data-based decisionmaking.

Data Governance: Changing Culture, Breaking Down Silos, and Deciding Who is in Control

Data Quality Campaign (DQC 2008).

http://dataqualitycampaign.org/files/meetings-dqc_quarterly_issue_brief-072908.pdf

This issue brief discusses the benefits of data governance, conditions helpful in successfully establishing data governance, and more. It also includes case studies of three states' experiences.

Governance Structure, LDS Project Team Organization, and Sustainability

http://nces.ed.gov/programs/slds/zip/granteemeeting06_5c.zip

In this presentation, several states report on the governance structures that have helped them succeed in implementing changes required in developing a longitudinal data system.

LDS Share

The following documents are available at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx:*

Effective Governance Models for Managing Data Systems

Tennessee Department of Education (2008).

This presentation details Tennessee's strategies and recommendations for implementing data governance.

Data Managers Working Group Charter

Michigan Department of Education (2007).

This document outlines the key roles in Michigan's Data Managers Working Group, which is comparable to the Data Governance Committee in this guide. Also included are the group's mission and its members' responsibilities.

Chapters 4–5: Data Quality

Forum Guide to Building a Culture of Quality Data: A School and District Resource

National Forum on Education Statistics (2004).

http://nces.ed.gov/forum/pub_2005801.asp

This guide was developed to help schools and school districts improve the quality of data they collect, and to provide processes for developing a "culture" of quality data by focusing on data entry–getting things right at the source. This guide shows how quality data can be achieved through the collaborative efforts of all staff.

Forum Curriculum for Improving Education Data: A Resource for Local Education Agencies

National Forum on Education Statistics (2007).

http://nces.ed.gov/forum/pub_2007808.asp

This curriculum supports efforts to improve the quality of education data by serving as training materials for K–12 school and district staff. It provides lesson plans, instructional handouts, and related resources; and presents concepts necessary to help schools develop a culture for improving data quality.

Texas Education Agency Data Standards

http://ritter.tea.state.tx.us/peims/standards/0910/index.html

This web page presents up-to-date documentation on the Texas data standards and requirements. Section 1 includes a discussion of data-submission training.

Texas Data Validation Monitoring 2008–2009

http://ritter.tea.state.tx.us/pmi/datamon/2009/index.html

This web page includes resources related to the Texas Education Agency's data validation system, which analyzes data submitted by districts and charters and automatically identifies potential data anomalies in their reporting on leavers and dropouts, student assessments, and discipline.

Kansas Data Quality Certification Program
Kansas State Department of Education.

http://www.ksde.org/default.aspx?tabid=3377

This website includes descriptions of the tracks for various roles (e.g., data entry personnel, administrators, etc.), many helpful resources, and more. Also see a presentation about the program, *Yellow Brick by Yellow Brick: Using a Professional Development Program to Strengthen Data Quality in Kansas,* at *http://nces.ed.gov/whatsnew/conferences/mis/2009/presentations/v_d.zip.*

Nebraska's Data Validation Process
Beecham and Tagart (2008).

http://nces.ed.gov/programs/slds/zip/08_f_10a.zip

This presentation, and the supporting document, outlines Nebraska's approach to data validation, which flags errors and presents data issues through reports on a versatile website. There, authorized staff members can review potential errors, make corrections, and approve the information.

The Process for Ensuring Data Quality
ESP Solutions (2009).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Process_for_Ensuring_Data_Quality_ORB%282%29.pdf

This two-part series includes best practices and principles to help education agencies achieve better data quality.

Process Illustration: Steps for Ensuring Data Quality
ESP Solutions (2004).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Data_Quality_Illustration.pdf

This illustration details key questions and steps involved in ensuring data quality.

The Data Quality Imperative, Data Quality Series—Part I
ESP Solutions (2007).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Data_Quality_Imperative_ORG.pdf

This is the first part of a series that merges foundations of data quality from the formal literature on information systems with practical lessons on ensuring data quality in the public education system.

The Data Quality Manual, Data Quality Series—Part II
ESP Solutions (2007).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Data_Quality_Manual_PartII_ORG.pdf

This is the second part of a series on data quality. It focuses on practical lessons learned on ensuring data quality gained from real school experiences.

LDS Share

The following documents are available at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx:*

Policies and Procedures: Verification of Electronic Data Systems and Accompanying Letter

State of Tennessee Department of Education (2006).

This document establishes policies and procedures for governing data system verification of quality. The accompanying letter is also available.

Statistical Process Control

Maryland State Department of Education (2007).

This document describes Maryland's statistical process control, a validation process added to the state's education data collection system to help avoid errors published online for state and federal compliance reporting.

Chapter 6: Data Standards

NCES Handbooks Online

National Center for Education Statistics.

http://nces.ed.gov/programs/handbook

The NCES *Data Handbooks* provide guidance on consistency in data definitions and maintenance for education data, so that such data can be accurately aggregated and analyzed. The online database provides the nonfiscal handbooks in a searchable web tool and includes data elements for students, staff, and education institutions.

National Education Data Model

http://nces.ed.gov/forum/datamodel

The National Education Data Model catalogs the data used in P–12 education and describes the relationships among those data. It is designed to be used as a reference tool to facilitate the identification, merging, and matching of data across different systems; to provide similar descriptions across local education agency (LEA) systems, across LEAs, and from LEAs to the state and federal government; and to specify the content and structure of logical and physical data models.

Common Education Data Standards Initiative

http://commondatastandards.org

This initiative is a national, collaborative effort to develop voluntary, common data standards for a key subset of K–12 (e.g., demographics, program participation, course information) and to-postsecondary education transition variables. The initiative aims to identify a list of key K–12 and to-postsecondary transition variables; and agree on standard definitions, code sets, business rules, and technical specifications for these variables. Expansion into PK and the workforce will be considered in the future.

DQC Standards, Interoperability, and Portability Thought Leaders Meeting

Data Quality Campaign (DQC 2009).

http://www.dataqualitycampaign.org/events/247

This webpage includes a number of presentations from a 2009 policymakers meeting in Washington, DC, as well as resources related to data standards.

SIF Association Implementation Toolkit

Schools Interoperability Framework (SIF) Association.

http://www.sifinfo.org/us/tool_kit.asp

This collection of documents is intended to help education institutions in the SIF implementation process. It includes planning questions (scope, desired automation, data needs, expected changes), RFP language, an implementation planning toolkit, and support resources.

Financial Accounting for Local and State School Systems: 2003 Edition

National Center for Education Statistics (2004).

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2004318

This NCES handbook was designed to be used as the national standard for state education departments (SEA) reporting financial data, and for school districts preparing comprehensive annual financial reports (CAFR).

The Right Data to the Right People at the Right Time: How Interoperability Helps America's Students Succeed

Data Quality Campaign (DQC 2007).

http://dataqualitycampaign.org/files/Meetings-DQC_Quarterly_Issue_Brief_061307.pdf

This paper reviews the needs for, benefits of, and concurrent efforts to establish interoperable education systems. It offers several key definitions, a case study section, and a list of interoperability examples from other industries.

Secondary School Course Classification System: School Codes for the Exchange of Data (SCED)

National Center for Education Statistics (2007).

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007341

This NCES handbook provides a taxonomy for assigning standard codes to secondary school courses in 22 major subject areas. It also includes a content description for each course, and instructions on how to use the taxonomy in coding courses.

Articulating the Case for Course Numbers

ESP Solutions (2006).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Course_Numbers_ORG.pdf

This booklet argues for the use of common course codes nationwide.

The Open Group

http://www.opengroup.org

The Open Group is a consortium of public and private sector organizations that promote open standards and global interoperability in order to enable access to integrated information both within and between enterprises. The organization provides a SIF certification program for vendors.

Metadata and data dictionaries

Forum Guide to Metadata: The Meaning Behind Education Data

National Forum on Education Statistics (2009).

http://nces.ed.gov/forum/pub_2009805.asp

This guide empowers more effective use of data as information. To accomplish this, the publication explains what metadata are, why they are critical to sound education data systems, what components comprise a metadata system, what value metadata bring to data management and use, and how to implement and use a metadata system.

Minnesota Department of Education Data Dictionary

Minnesota Department of Education.

http://education.state.mn.us/mde-dd

Minnesota has prepared and published its education data dictionary. This version has the same content and form as the state's internal data dictionary, except that data elements that might be sensitive from a security point of view are not revealed.

South Carolina LDS Project: Data Dictionary–Data Model

South Carolina Department of Education (2006).

Available via LDS Share at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx*

This document illustrates South Carolina's longitudinal data system project's data dictionary/data model.

LDS Share

The following documents are available at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx:*

Strategic Approach For Developing the Longitudinal Data System Data Dictionary

Wisconsin Department of Public Instruction (2007).

This working document outlines the approach for developing a data dictionary for the Wisconsin Department of Public Instruction's longitudinal data system.

Agency-Wide Data Dictionary Planning Project

Wisconsin Department of Public Instruction (2007).

This document details the planning and the implementation projects related to the Wisconsin Department of Public Instruction's agency-wide data dictionary.

Chapter 7: Privacy and Confidentiality

Basic Concepts and Definitions for Privacy and Confidentiality in Student Education Records

National Center for Education Statistics (2010).

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011601

This technical brief discusses basic concepts and definitions that establish a common set of terms related to the protection of personally identifiable information, especially in education records in the Statewide Longitudinal Data Systems (SLDS). This Brief also outlines a privacy framework that is tied to Fair Information Practice Principles that have been promulgated in both the United States and international privacy work.

Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting

National Center for Education Statistics (2010).

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011603

This technical brief examines what protecting student privacy means in a reporting context. To protect a student's privacy, the student's personally identifiable information must be protected from public release. When schools, districts, or states publish reports on students' educational progress, they typically release aggregated data—data for groups of students—to prevent disclosure of information about an individual. However, even with aggregation, unintended disclosures of personally identifiable information may occur. Current reporting practices are described and each is accompanied by an example table that is used to consider whether the intended protections are successful.

Data Stewardship: Managing Personally Identifiable Information in Student Education Records

National Center for Education Statistics (2010).

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011602

This technical brief focuses on data stewardship, which involves each organization's commitment to ensuring that privacy, confidentiality, security, and the appropriate use of data are respected when personally identifiable information is collected. Data stewardship involves all aspects of data collection, from planning, collection and maintenance to use and dissemination. The Brief also discusses internal control procedures that should be implemented to protect personally identifiable information, including the use of unique student identifiers and linking codes, workforce security, authorization for access, role based access to student record data, permitted uses, and the handling of data breaches. This Brief concludes with a discussion of accountability and auditing, including an overview of the types of audit activities that can be implemented to ensure that all stages of data stewardship have been successfully implemented.

Forum Guide to the Privacy of Student Information: A Resource for Schools

National Forum on Education Statistics (2006).

http://nces.ed.gov/forum/pub_2006805.asp

This guide helps school and local education agency staff to better understand and apply FERPA, a federal law that protects privacy interests of parents and students in education

records. (Note: The Forum is currently developing a single revised guide that will replace all current privacy resources about schools, LEAs and SEAs.)

Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies
National Forum on Education Statistics (2004).

http://nces.ed.gov/forum/pub_2004330.asp

This guide presents a general overview of privacy laws and professional practices that apply to information collected for, and maintained in, student records. The document also provides an overview of key principles and concepts governing student privacy; summarizes Federal privacy laws including recent changes; identifies issues concerning the release of information to both parents and external organizations; and suggests good data management practices for schools, districts, and state education agencies. (Note: The Forum is currently developing a single revised guide that will replace all current privacy resources about schools, LEAs and SEAs.)

Privacy Issues in Education Staff Records
National Forum on Education Statistics (2000).

http://nces.ed.gov/forum/pub_2000363.asp

This publication addresses key concepts in protecting and managing information on staff records. It does not provide legal guidelines but, rather, addresses the federal Freedom of Information and Privacy Acts, and offers principles of best practice.

Redefining Student Data Access Policy
Legislative Analyst's Office (2008).

http://www.lao.ca.gov/2008/edu/student_data_access/student_data_access.pdf

This brief proposes policy changes to the California legislature that would ease limitations on instructors' and policymakers' use of student data to improve instruction while still protecting student privacy.

Protecting Student Records and Facilitating Education Research: A Workshop Summary
National Research Council (2008).

http://www.nap.edu/catalog.php?record_id=12514#toc

To explore possibilities for data access and confidentiality in compliance with FERPA, the National Academies and the American Educational Research Association convened the Workshop on Protecting Student Records and Facilitating Education Research in April 2008. This document contains a wealth of information on a broad range of related subjects.

Confidentiality and Reliability Rules for Reporting Education Data
ESP Solutions (2008).

http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Confidentiality_Reliability_ORG.pdf

This extensive report fleshes out the confidentiality and reliability issues that states encounter when reporting education data. It also offers criteria for agencies to use when establishing the rules for, and selecting, a minimum number of students to report as subgroups in data sets; and performance reports such as adequate yearly progress (AYP). In particular, see pages 26–32 for a discussion on methods of manipulating data to protect confidentiality.

Family Educational Rights and Privacy Act (FERPA)

Family Policy and Compliance Office (FPCO)

http://www2.ed.gov/policy/gen/guid/fpco/index.html

This office of the U.S. Department of Education administers FERPA, and its website contains a host of resources on the law including current regulations, legislative history, and guidance on disclosures.

Final FERPA Regulations

U.S. Department of Education (2008).

http://www2.ed.gov/legislation/FedRegister/finrule/2008-4/120908a.pdf

These regulations were issued by the U.S. Department of Education and have been effective since January 2009. This revised interpretation of the law contains extensive discussions of public comments received about the previous version of the regulations, and all changes are noted.

FERPA Final Rule: Section-by-Section Analysis

Family Policy and Compliance Office (FPCO 2008).

http://www2.ed.gov/policy/gen/guid/fpco/pdf/ht12-17-08-att.pdf

This document highlights the changes in the U.S. Department of Education's new interpretation of FERPA.

Forum Guide to the Privacy of Student Information: A Resource for Schools

National Forum on Education Statistics (2006).

http://nces.ed.gov/forum/pub_2006805.asp

This guide helps school and local education agency staff to better understand and apply FERPA, a federal law that protects privacy interests of parents and students in education records. (Note: The Forum is currently developing a single revised guide that will replace all current privacy resources about schools, LEAs and SEAs.)

Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies

National Forum on Education Statistics (2004).

http://nces.ed.gov/forum/pub_2004330.asp

This guide presents a general overview of privacy laws and professional practices that apply to information collected for, and maintained in, student records. The document also provides an overview of key principles and concepts governing student privacy; summarizes Federal privacy laws including recent changes; identifies issues concerning the release of information to both parents and external organizations; and suggests good data management practices for schools, districts, and state education agencies. (Note: The Forum is currently developing a single revised guide that will replace all current privacy resources about schools, LEAs and SEAs.)

FERPA Resource Page
Data Quality Campaign.

http://dataqualitycampaign.org/resources/topics/13

This page contains a wealth of resources on FERPA, including policy briefs, legal analyses, state memoranda of agreement and understanding, and more.

FERPA Final Regulations and Analysis 2008 Resource Page
Data Quality Campaign (DQC).

http://dataqualitycampaign.org/resources/topics/14

This page contains several helpful resources on the 2008 FERPA amendments, including analysis; suggestions for further revisions; and FERPA Mythbusters, a document that addresses some misconceptions about the privacy law.

Maximizing the Power of Education Data While Ensuring Compliance with Federal Student Privacy Laws: A Guide for State Policymakers
Data Quality Campaign (DQC 2007).

http://dataqualitycampaign.org/files/Publications-FERPA_A_Guide_for_State_Policymakers.PDF

This issue brief analyzes how the new roles of state education agencies (and their longitudinal data systems) in data collection and sharing can be aligned with FERPA. Though this document is based on a particular interpretation of FERPA, it may serve as a guide to states as they build and use longitudinal data systems in ways that comply with FERPA and fully protect the privacy rights of students and parents.

Health Insurance Portability and Accountability Act (HIPAA)

Department of Health and Human Services (HHS) Health Information Privacy
http://www.hhs.gov/ocr/privacy/index.html

This website provides fact sheets, educational materials, and frequently asked questions about HIPAA.

Joint Guidance on the Application of FERPA and HIPAA To Student Health Records
Family Policy Compliance Office (FPCO 2008).

http://www2.ed.gov/policy/gen/guid/fpco/doc/ferpa-hippa-guidance.pdf

The document explains the relationship between FERPA and the HIPAA Privacy Rule, and addresses any confusion on the part of school administrators, health care professionals, and others as to how these two laws apply to student records.

Issue Brief on Privacy Standards for Student Health Records
National Association of School Nurses (2004).

http://www.nasn.org/default.aspx?tabid=277

Geared towards school health care providers, this resource discusses the relationship between FERPA and HIPAA and the laws' implications on their work. The association also has a resource page on FERPA and HIPAA: *http://www.nasn.org/default.aspx?tabid=146.*

Special Education Technical Assistance Paper

Washington State Office of Public Instruction (2003).

http://www.k12.wa.us/SpecialEd/pubdocs/tap4.pdf

This brief paper discusses the instances under which educational institutions are subject to HIPAA.

Forum Curriculum for Improving Education Data: A Resource for Local Education Agencies

National Forum on Education Statistics (2007).

http://nces.ed.gov/forum/pub_2007808.asp

Page 41 of this resource discusses privacy requirements for health records, including HIPAA and its interaction with FERPA.

Individuals with Disabilities Education Act (IDEA)

Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies

National Forum on Education Statistics (2004).

http://nces.ed.gov/forum/pub_2004330.asp

See pages 17–18 of this guide for a discussion of privacy protection under IDEA, and the law's relationship to FERPA. (Note: The Forum is currently developing a single revised privacy product that will replace all current privacy resources about schools, LEAs, and SEAs.)

Building the Legacy: IDEA 2004

U.S. Department of Education.

http://idea.ed.gov/explore/home

This site offers a large collection of resources related to Part B of IDEA, which deals with children aged 3–21.

Office of Special Education and Rehabilitative Services

U.S. Department of Education.

http://www2.ed.gov/about/offices/list/osers/osep/policy.html

This site offers a host of resources related to IDEA legislation, regulations, and policy documents intended to help states implement the law.

Joint Guidance on the Application of FERPA and HIPAA to Student Records

Family Policy Compliance Office (FPCO 2008).

http://www2.ed.gov/policy/gen/guid/fpco/doc/ferpa-hippa-guidance.pdf

While this document focuses mainly on FERPA and HIPAA, a brief discussion of IDEA is offered in section II.

National School Lunch Act (NSLA)

National School Lunch Program

U.S. Department of Agriculture, Food, and Nutrition Service.

http://www.fns.usda.gov/cnd/lunch

This site contains many resources on NSLA, including a fact sheet and regulations.

Confidentiality of Free and Reduced Price Meal Eligibility

Arkansas Department of Education (2009).

http://arkedu.state.ar.us/commemos/static/fy0809/4093.html

This memo discusses privacy requirements under NSLA, and how they relate to FERPA.

Chapter 8: Data Security

Enterprise Data Security

Birrittieri, T. (2008). *InfoManagement Direct,* November 2008.

http://www.information-management.com/infodirect/2008_99/10002235-1.html

This article offers a simple "ownership" model of data security, focusing specifically on identifying data sensitivity and appropriately granting access to data user groups. It argues that group delineations should be based on users' job functions rather than the data's level of confidentiality.

Forum Unified Education Technology Suite

National Forum on Education Statistics (2005).

http://nces.ed.gov/forum/pub_tech_suite.asp

Part 5 of this resource addresses security, including tips on risk assessment and policy development.

Objective of the Data Security Model

Zhao, X.; O'Connor, B.; and Barraso, G. (2006). *InfoManagement Direct,* August 2006.

http://www.information-management.com/infodirect/20060825/1061606-1.html

This article provides a technical discussion of the data security "ownership" model. It describes a suggested set of processes including data classification, the identification of the risks associated with business functions and data systems, the creation and implementation of an access control policy, and the establishment of training and contingency plans.

Data Environment Security

Arizona Department of Education (2008).

http://nces.ed.gov/forum/zip/sum08_techsecurity1.zip

This presentation provides an overview of Arizona's efforts to improve security and manage data access. It discusses the establishment of an office to focus on security challenges, as well as creating processes for managing and monitoring data usage, training staff, and developing a means for communicating security issues. The presentation also provides an overview of some of the technical solutions the agency has employed.

Restricted-Use Data Procedures Manual

Institute of Education Statistics (2007).

http://nces.ed.gov/statprog/rudman

This IES manual provides a guide to the restricted-use data application process, and to the laws and regulations governing these data.

Managing Data Systems Security

http://nces.ed.gov/whatsnew/conferences/mis/2009/session_v.asp#a

New Hampshire Department of Education (2009).

This presentation briefly introduces New Hampshire's education agency's single sign-on system for user-identity management.

LDS Share

The following documents are available at *http://nces.ed.gov/programs/slds/ldsshare/slds*.aspx:

Compiled Data Access and Use Agreements

Virginia Department of Education (2006).

This document from 2006 compiles data access and use agreements from Kansas, Oregon, Oklahoma, Illinois, Missouri, and Louisiana.

Policies and Procedures for Data Security and Confidentiality

Tennessee Department of Education (2005).

This sample shows how a state education agency documented policies and specific supporting procedures to guide governance related to data security and confidentiality.

Memorandum of Agreement Between Alaska Department of Education and Early Development and Computer Task Group

Alaska Department of Education and Early Development (2006).

Available via LDS Share at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx*

This sample memorandum of agreement authorized a vendor's access to education data maintained and collected by the Alaska Department of Education and Early Development for the purpose of carrying out its contractual obligations related to the delivery of a portion of the agency's data dictionary project.
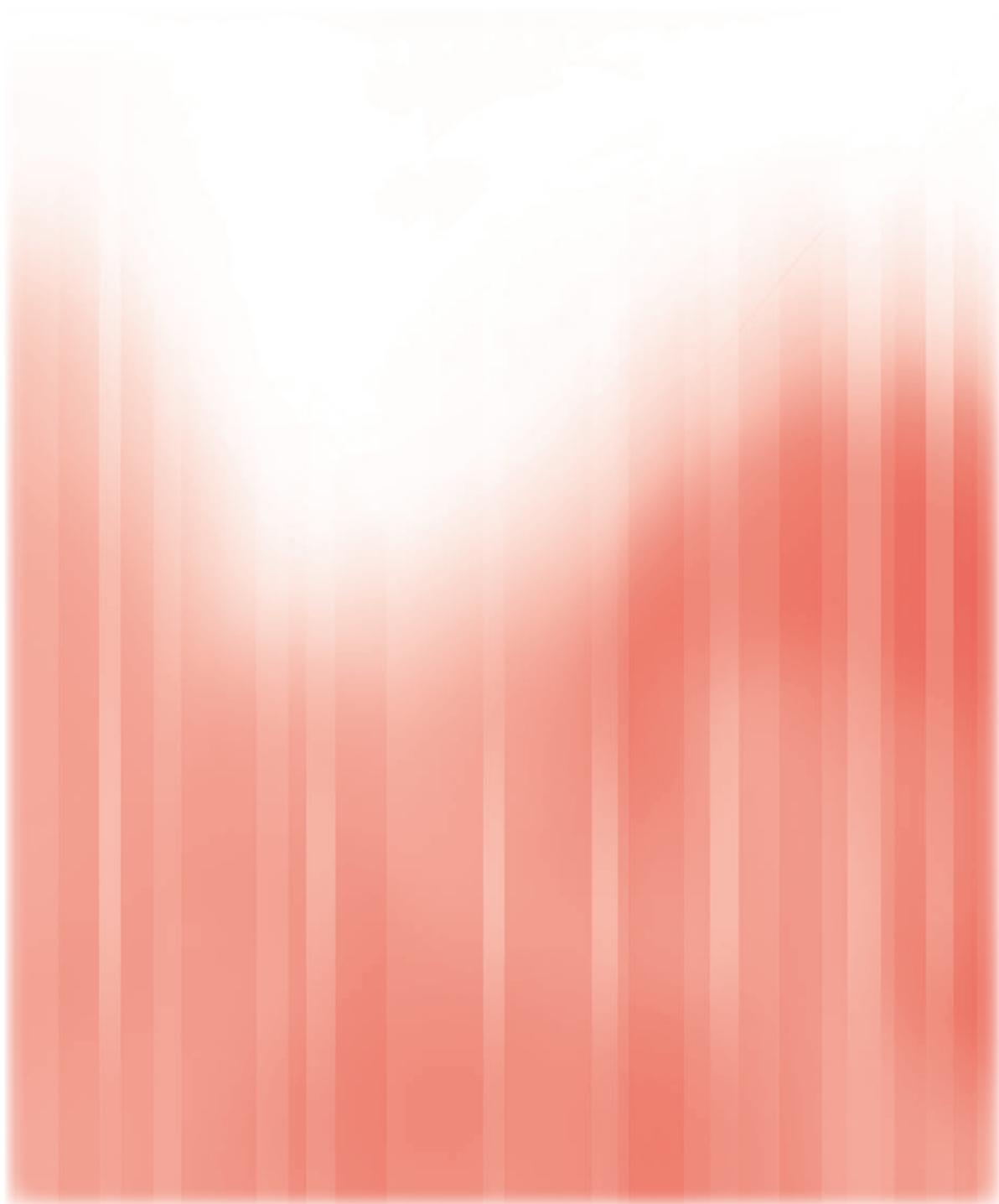
Handout: The Ungoverned Education Agency

The History and Consequences

The state education agency's data collection and management practices had come about over time, driven mainly by compliance and funding. Various program areas were created to focus on specific federal surveys, and staff collected the data needed to do their jobs. Program area staff administered the surveys, followed their own quality assurance processes, and maintained and secured the data in their own silo systems. And, of course, data were reported as required by the federal government. Individual managers took their own approach to directing staff and organizing work, and coordination across program areas was limited.

Over time, different departments began to collect some of the same data elements. Inconsistencies were commonplace. For instance, while the student information system listed an "Aileen Hutchinson," who was not in special education, the Special Education Department's system included a girl named "Allie N Hutchinsen." Despite these discrepancies, the staff knew the two records referred to the same student based on other directory information. However, structural differences between the systems–different definitions, formats, and option sets–further complicated matters. For example, Allie's race, White, was coded as "1" in one database and "2" in the other because the systems used different option sets. Furthermore, because program areas defined their own data elements and used different software to manage them, the ability of the agency's many data systems to "talk" to one another varied from limited to nonexistent, burdening staff with redundant data entry work and introducing errors into the system. With no clear requirement for documentation of data processes, methodologies often changed, as happened when Joe left the agency and no one else knew how to produce the dropout rate. The new guy, Steve, calculated it the way he had at his previous agency. No one saw a problem with this, especially since Steve's numbers were lower than Joe's. But when the time came to compare the new rates to previous year's, some staff realized that they were comparing apples to oranges.
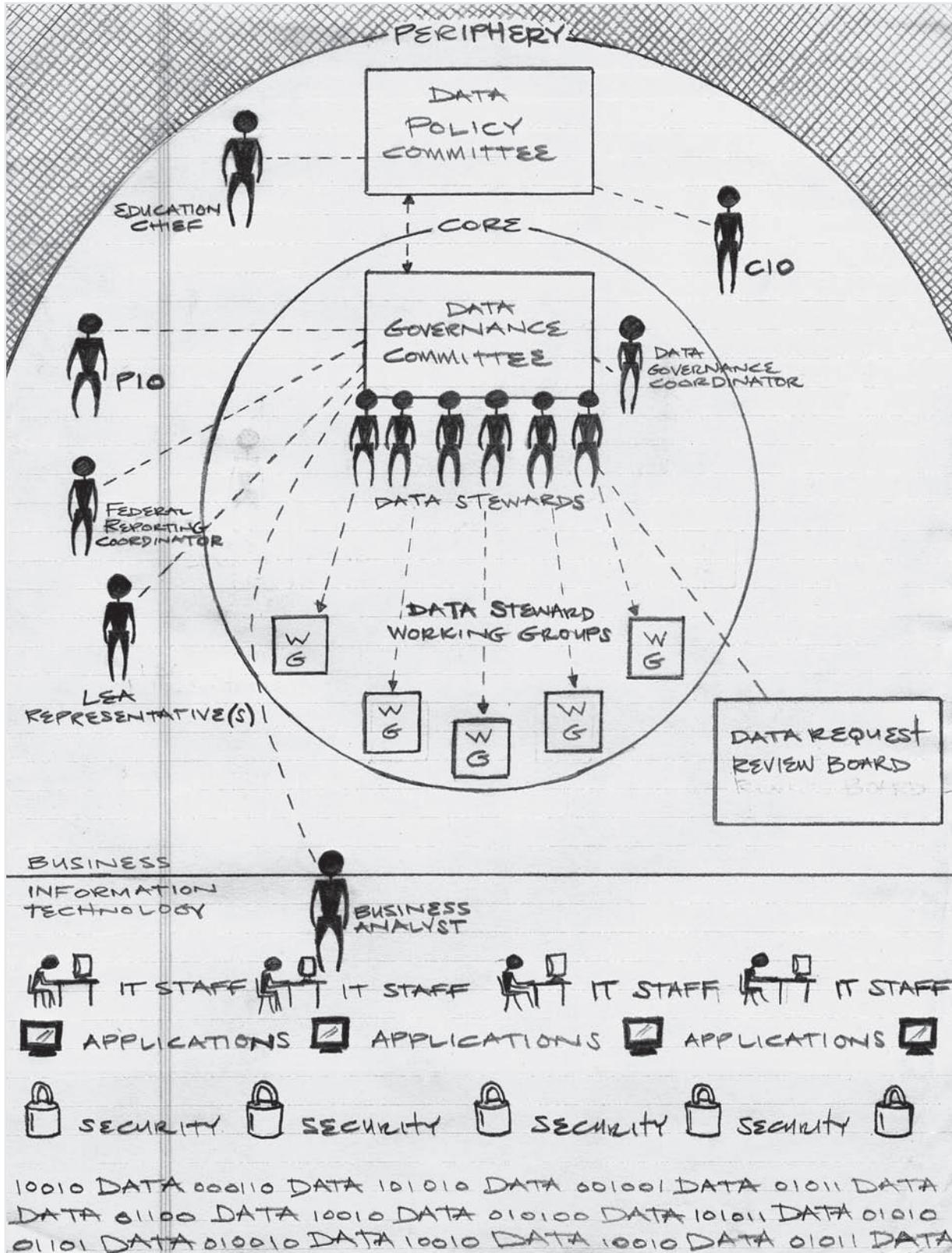
Year in and year out, the work was done, but specific tasks weren't assigned to anyone in a consistent manner, and sometimes reports were late or incomplete. And since there was no "official" source for each data element, data for federal reports might come from one source one year, another the following year. Similarly, without clear guidelines, staff fielded data requests as best they could. When inquiries came in, the recipient would decide to which program area the request should be sent. In one instance, Talia was asked for the school addresses in a district. While McKenna and Vita both managed school directory data, in separate systems, Vita was given the task because her desk was near Talia's.

When postsecondary education leaders asked about sharing data, agency officials cried "FERPA!," invoking the federal privacy law they mistakenly thought prohibited such exchange (see chapter 7). Security and access protocols were not well understood, and staff often took a lax approach to protecting sensitive information (see "Identity theft in the printer room" in chapter 8). Data quality problems were handled as they arose, but no long-term changes were made to ensure the same issues wouldn't crop up again. And since no responsible group or process was in place to identify the sources of the agency's problems, they went undiscovered and data quality issues were just blamed on IT. This went on for some time and, other than the techies, most staff didn't see much of a problem. It was simply business as usual.

The Realization

Over time, the country became more interested in education data. Education stakeholders wanted more information for accountability purposes and to better understand what programs and instructional strategies worked. They wanted data to inform decisionmaking at all levels; and to improve administration, instruction, and student performance. The bottom line was, they wanted data from across the agency and they wanted them fast. This changing environment posed many problems for the agency. Requested analyses required linking across silos, or data integration in a central data store. Before this could happen, duplicate, inconsistent data had to be reconciled. However, once the integration work began, more inconsistencies were discovered than anyone had imagined. Data quality had to be a higher priority, security had to be improved, and the data elements collected had to serve business and stakeholder needs, not just meet federal requirements.

Better methods of sharing data had to be devised if the agency was to meet the growing demand for a "P–20" system. And better, more consistent protocols were needed to make data sharing more efficient and prevent improper dissemination. The chief information officer decided something had to be done. Having seen a presentation at a national conference, he was convinced a process called "data governance" could help address the agency's problems.

Appendix E

Data Governance Supplementary Materials

Sample Data Governance Committee Mission Statement, Goals, and Objectives

(Adapted with permission from Chatis Consulting)

Example Data Governance Committee Mission

The Data Governance Committee supports the _____ Department of Education's mission of helping teachers teach and children learn by promoting the appropriate use of data to inform decisionmaking; and ensuring data quality, accountability, and timeliness.

Examples of Data Governance Committee Goals

➤ Improve data quality.

➤ Increase accountability for data accuracy.

➤ Eliminate redundancy in data collection.

➤ Improve understanding of data within the department and among districts.

➤ Increase data use in program and policy decisions.

➤ Improve capability and timeliness of data reporting.

Examples of Data Governance Committee Objectives

➤ Identify the owner of every data element.

➤ Define all data elements.

➤ Document all data processes.

➤ Standardize data processes from year to year.

➤ Reduce manual manipulation of data.

➤ Identify the official source of data for all external reporting.

➤ Eliminate redundant data collections that are not the official source for external reporting.

➤ Allow districts to review their data before they are reported externally.

➤ Communicate all data decisions/changes to districts.

➤ Reduce the collection of, and reliance on, aggregate data.

➤ Increase the use of student-level data for external reporting.

Sample Data Stewards Working Group Guidelines

(Adapted with permission from Chatis Consulting)

Process for addressing critical data issues

1. Identify the data steward responsible for the issue and its resolution (*one* person).

2. Determine whether a small working group of relevant data stewards should be created to address the issue.

   a. Does the issue *directly* affect the data quality or work of more than one program/subject area in the organization?

   b. Note: Even if a working group is formed, only *one* data steward should be accountable for the issue.

3. Plan first meeting of the working group.

   a. Clearly define (and document) the *source* of the problem–not the symptoms. This includes all aspects of the issue: communication (internal and external), definitions, technology, etc. A reporting problem is almost never *just* a reporting problem: its source is earlier in the data process. (Note: If the issue is complex, additional research and time could be require to fully identify it; this time spent at the beginning of the process is well worth it to fully understand what you are trying to address.)

   b. Determine the goals of addressing the problem; what exactly does the group want to achieve? (These goals should be aligned with the Data Governance Committee's goals).

4. Create a mini-project plan for addressing each aspect of the problem and for achieving the established goals.

   a. Include main steps, including due dates and who must be involved in, or responsible for, each.

   b. Assign action items at the end of each workgroup session, including responsible staff member and due date.

   c. Determine whether any part of the issue is not within the purview of the Data Governance Committee. If so, the responsible data steward should bring the issue to the Data Governance Committee chair.

   d. Provide monthly updates to the Data Governance Committee for inclusion on the critical-data issues log.

5. Once a preliminary "business" solution has been developed (i.e., you know what you want to do), coordinate with technology staff to get their input and determine how it can be implemented.

6. Document all decisions made and implemented solutions thoroughly, and save to a common online area accessible by all Data Governance Committee members.

7. Communicate these decisions and implemented solutions to the Data Governance Committee, all applicable program areas, LEAs, and any other staff directly affected by the issue. Be especially clear if the solution requires certain staff members to change how they work.

8. Retire the issue from the critical data issues log, celebrate! Then move on to the next issue.

Sample Critical Data Issues Log

Data governance committee: Critical data issues

Criteria: Creates a burden for the districts, causes errors or delays in federal reporting, or prevents the use of the data for its intended purpose.

Data item	Issue description	Data stewards(s)	Priority	Recent action taken?	Status update	Action items

Retired data issues

Data item	Issue description	Data stewards(s)	Priority	Date retired	Last status update	Reason retired

Source: Adapted from the Tennessee Department of Education's Data Management Committee Critical Data Issues Log. Available via LDS Share at *http://nces.ed.gov/programs/slds/ldsshare/slds.aspx.*
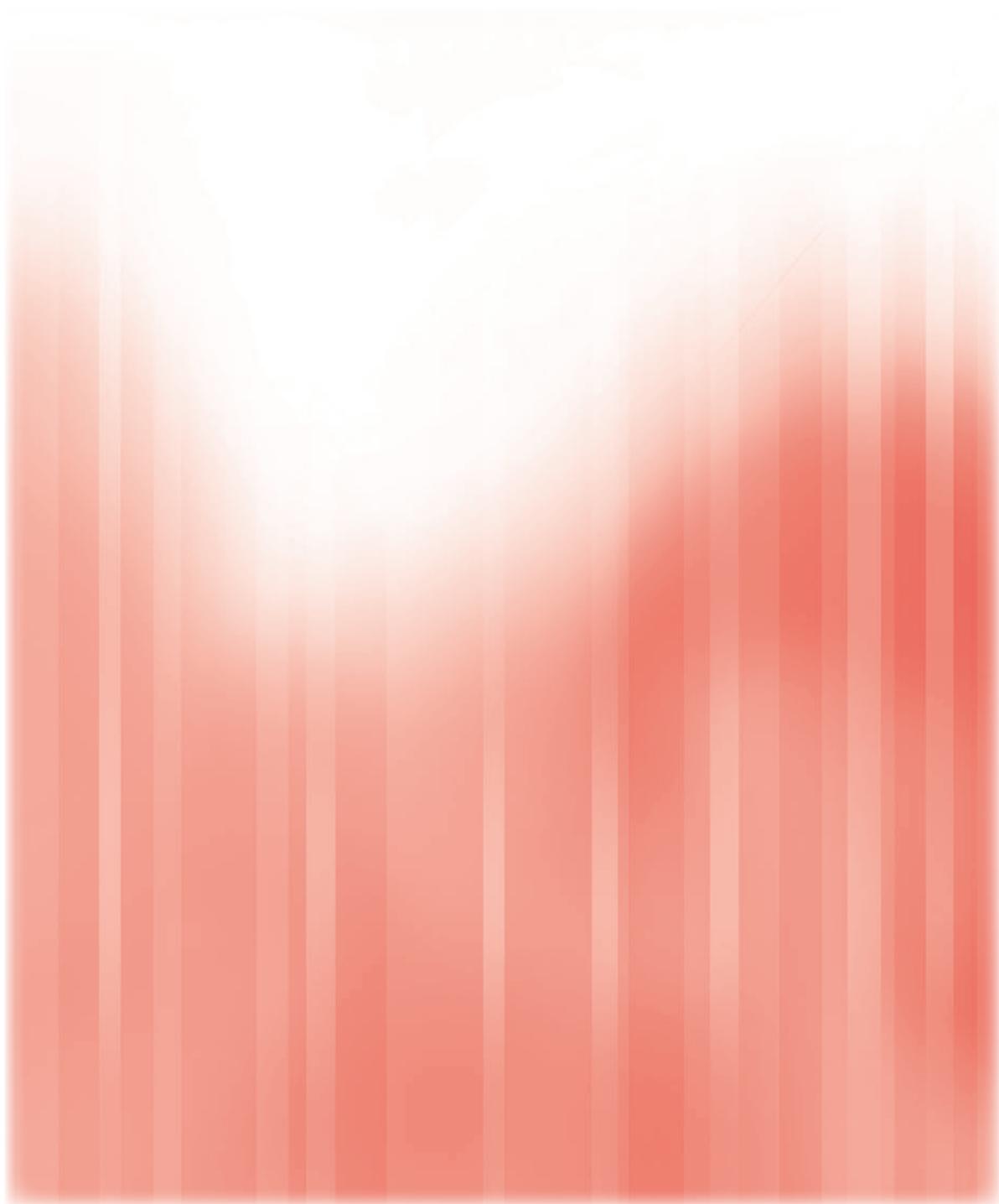
Appendix F

Forum and Other NCES Resources

The Forum Guide to Data Ethics (NFES 2010–801)
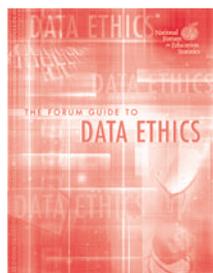
http://nces.ed.gov/forum/pub_2010801.asp

Every day, educators collect and use data about students, staff, and schools. Some of these data originate in individual student and staff records that are confidential or otherwise sensitive. And even those data that are a matter of public record, such as aggregate school enrollment, need to be accessed, presented, and used in an ethically responsible manner. While laws set the legal parameters that govern data use, ethics establish fundamental principles of "right and wrong" that are critical to the appropriate management and use of education data in the technology age. This guide reflects the experience and judgment of experienced data managers; while there is no mandate to follow these principles, the authors hope that the contents will prove a useful reference to others in their work.

Crisis Data Management: A Forum Guide to Collecting and Managing Data about Displaced Students (NFES 2010–804)

http://nces.ed.gov/forum/pub_2010804.asp

This publication provides guidelines that can be used by elementary and secondary education agencies to establish policies and procedures for collecting and managing education data before, during, and after a crisis.

The Forum Guide to Metadata: The Meaning Behind Education Data (NFES 2009–805)
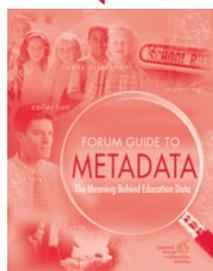
http://nces.ed.gov/forum/pub_2009805.asp

This guide offers best practice concepts, definitions, implementation strategies, and templates/tools for an audience of data, technology, and program staff in state and local education agencies. This resource was developed to improve these audiences' awareness and understanding of metadata and, subsequently, the quality of the data in the systems they maintain.

Every School Day Counts: Forum Guide to Collecting and Using Attendance Data (NFES 2009–804)
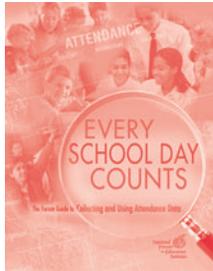
http://nces.ed.gov/forum/pub_2009804.asp

This guide offers best practice suggestions on collecting and using student attendance data to improve performance. It includes a standard set of codes to make attendance data comparable across districts and states. The publication includes real-life examples of how attendance information has been used by school districts.

NCES Handbooks and NCES Handbooks Online

http://www.nces.ed.gov/programs/handbook

The NCES *Handbooks* are a valuable source of metadata for organizations and individuals interested in education data. These print and online resources define standard education terms for students, staff, schools, local education agencies (LEA), intermediate education agencies, and state education agencies (SEA). The Handbooks are intended as reference documents for public and private organizations, including education institutions and early childhood centers; as well as education researchers and other users of education data. In order to improve access to this valuable resource, NCES has also developed the NCES *Handbooks Online,* a web-based tool that allows users to view and download information via an electronic table of contents, a drill-down finder, element name and first letter searches, and advanced query options.

National Education Data Model

http://nces.ed.gov/forum/datamodel/index.aspx

The National Education Data Model (NEDM) is the first non-proprietary, national education data model developed to help schools, LEAs, and states design or guide the selection of systems for instructional delivery, data-driven decisionmaking, data collection, operations, and reporting. The model provides a national blueprint to help schools evaluate and improve instructional tools; communicate needs to their umbrella agency or to vendors; enhance the movement of student information from one LEA to another; and, in the end, have better tools to inform instruction. NEDM can be used by educators, vendors, and researchers to understand the information required for teaching, learning, and administrative systems.

Managing an Identity Crisis: Forum Guide to Implementing New Federal Race and Ethnicity Categories (NFES 2008–802)
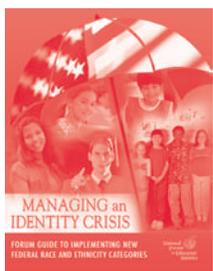
http://nces.ed.gov/forum/pub_2008802.asp

This best-practice guide was developed to help state and local education agencies implement the new federal race and ethnicity categories, thereby reducing redundant efforts within and across states, improving data comparability, and minimizing reporting burden. Users may select and adopt strategies that will help them quickly begin the process of implementation in their agencies.

Forum Guide to Core Finance Data Elements (NFES 2007–801)
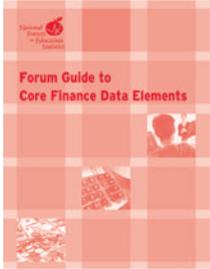
http://nces.ed.gov/forum/pub_2007801.asp

This publication establishes current and consistent terms and definitions for maintaining, collecting, reporting, and exchanging comparable information related to education finances. It is designed to accompany Financial Accounting for Local and State School Systems: 2003 Edition by identifying common reporting requirements and defining frequently used indicators and calculations that use data elements from accounting and other data systems.

Forum Curriculum for Improving Education Data: A Resource for Local Education Agencies (NFES 2007–808)

http://nces.ed.gov/forum/pub_2007808.asp

This resource supports efforts to improve the quality of education data by serving as training materials for K–12 school and district staff. It provides lesson plans, instructional handouts, and other resources; and presents concepts necessary to help schools develop a culture for improving data quality.

Forum Guide to Decision Support Systems: A Resource for Educators (NFES 2006–807)
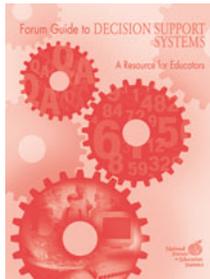
http://nces.ed.gov/forum/pub_2006807.asp

This Forum guide was developed to help the education community better understand what decision support systems are, how they are configured, how they operate, and how they might be developed and implemented in an education setting.

Forum Guide to Virtual Education (NFES 2006-803)

http://nces.ed.gov/forum/pub_2006803.asp

This publication offers recommendations for collecting accurate, comparable, and useful data about virtual education in elementary and secondary education settings. It highlights policy questions and data elements critical to meeting the information needs of policymakers, administrators, instructors, and parents of students involved in virtual education.

Forum Guide to the Privacy of Student Information: A Resource for Schools (NFES 2006–805)
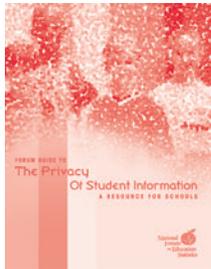
http://nces.ed.gov/forum/pub_2006805.asp

This publication was written to help school and local education agency staff better understand and apply the Family Educational Rights and Privacy Act (FERPA), a federal law that protects the privacy interests of parents and students with respect to information maintained in student education records. It defines terms such as "education records" and "directory information," and offers guidance for developing appropriate privacy policies and information disclosure procedures related to military recruiting, parental rights and annual notification, videotaping, online information, media releases, surveillance cameras, and health-related information.

Accounting for Every Student: A Taxonomy for Standard Student Exit Codes (NFES 2006–804)
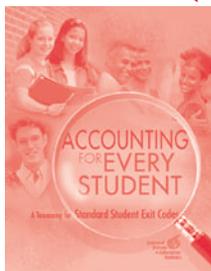
http://nces.ed.gov/forum/pub_2006804.asp

This publication was developed to help education agencies develop effective information systems for tracking students' enrollment status. It presents a student-level exit code taxonomy for states and districts that accounts for 100 percent (not 90 or 110 percent) of all students. It also offers "best practice" advice regarding tracking students, collecting exit codes data, and distinguishing among high school completion credentials.

Forum Guide to Education Indicators (NFES 2005–802)

http://nces.ed.gov/forum/pub_2005802.asp

This publication provides encyclopedia-type entries for 44 commonly used education indicators. Each indicator entry includes a definition, recommended uses, caveats and cautions, related policy questions, data element components, a formula, commonly reported subgroups, and display suggestions. The document will help readers better understand how to appropriately develop, apply, and interpret commonly used education indicators.

Forum Guide to Building a Culture of Quality Data (NFES 2005–801)

http://nces.ed.gov/forum/pub_2005801.asp

This publication focuses on data entry: getting things done right at the source. It recommends a practical process for developing a "Culture of Quality Data" based around individual tip sheets for individuals involved in providing data, including principals, teachers, office staff, school board members, superintendents, data stewards, and technology staff.

Forum Unified Education Technology Suite (2005)

http://nces.ed.gov/forum/pub_tech_suite.asp
This publication presents a practical, comprehensive, and tested approach to assessing, acquiring, instituting, managing, securing, and using technology in education settings. It is written for individuals without extensive experience with technology who have been tasked with leading technology initiatives in a school or district setting,.

Forum Guide to Protecting the Privacy of Student Information: State and Local Education Agencies (NCES 2004–330)

http://nces.ed.gov/forum/pub_2004330.asp
This publication presents a general overview of privacy laws and professional practices that apply to information collected for, and maintained in, student records. The guide provides an overview of key principles and concepts governing student privacy; summarizes federal privacy laws; identifies issues concerning the release of information to both parents and external organizations; and suggests good data management practices for schools, districts, and state education agencies.

Facilities Information Management: A Guide for State and Local Education Agencies (NCES 2003–400)

http://nces.ed.gov/forum/pub_2003400.asp
This publication provides a framework for identifying a basic set of school facilities data elements and definitions that will meet the information needs of school and community decisionmakers, school facility managers, and the general public. It presents recommendations for designing and maintaining an information system that addresses the condition, design, use, management, and financing of elementary and secondary education facilities. Commonly used measures, data elements, and additional resources for the practitioner are also included.

Planning Guide for Maintaining School Facilities (NCES 2003–347)

http://nces.ed.gov/forum/pub_2003347.asp
This publication is intended to help school facilities managers plan for efficient and effective operations. It provides practical advice on a range of topics, including how to conduct a facilities audit, plan for maintenance to ensure smooth operations and avoid costly surprises, manage staff and contractors, and evaluate maintenance efforts.