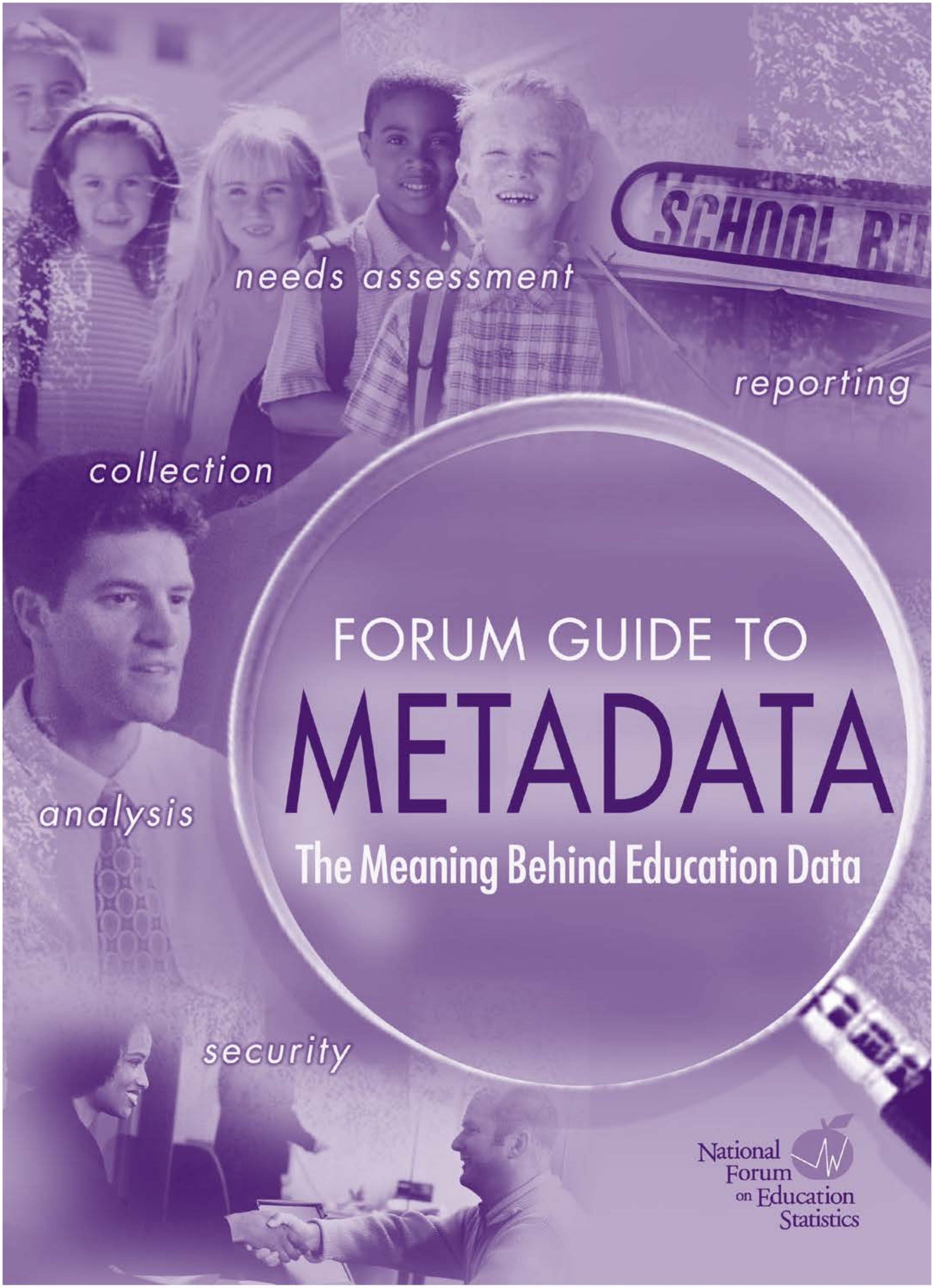


FORUM GUIDE TO METADATA

The Meaning Behind Education Data

National Forum on Education Statistics

National Cooperative Education Statistics System

The National Center for Education Statistics (NCES) 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 on 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) is an entity of the Cooperative System and, among its 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 NCES.

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

July 2009

This publication and other publications of the National Forum on Education Statistics may be found at the websites listed below.

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*

Suggested Citation
National Forum on Education Statistics. (2009). *Forum Guide to Metadata: The Meaning Behind Education Data* (NFES 2009–805). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

For ordering information on this report, write:
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

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

Technical Contact:
Ghedam Bairu
(202) 502-7304
Ghedam.Bairu@ed.gov

Task Force Members

This document was developed through the National Cooperative Education Statistics System and funded by the National Center for Education Statistics (NCES) of the U.S. Department of Education. A volunteer task force of the National Forum on Education Statistics produced this document.

Chair

Tom Ogle
Missouri Department of Elementary and Secondary Education

Task Force Members

Kathy Gosa
Kansas State Department of Education

Sonya Edwards
California Department of Education

Tom Saka
Hawaii Department of Education

Steve Smith
Cambridge Public Schools (Massachusetts)

Mike Strozeski
formerly of Richardson Independent SD (Texas)

Consultant

Tom Szuba
Quality Information Partners

Project Officer

Ghedam Bairu
National Center for Education Statistics

Acknowledgments

The members of the Metadata Task Force of the National Forum on Education Statistics would like to thank everyone who reviewed drafts of this document or otherwise contributed to its development. This includes the Forum Steering Committee, the Forum Technology (TECH) Committee, members of the National Forum on Education Statistics, and Lee Hoffman and Richard Struense of the National Center for Education Statistics. Additionally, the task force wishes to thank Tim Webb of the Tennessee Department of Education; and Rebecca Kaye, formerly of the U.S. Department of Education. Frances Erlebacher edited the document and the Creative Shop provided layout and design services.

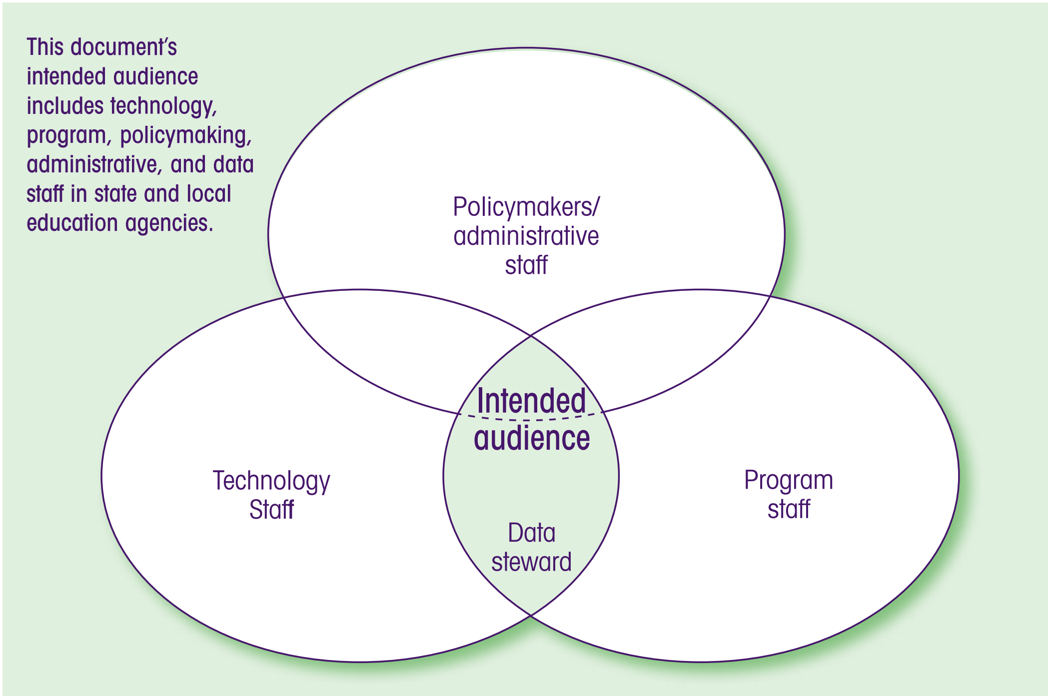
Foreword

The National Forum on Education Statistics (the Forum) is pleased to present the *Forum Guide to Metadata: The Meaning Behind Education Data*. 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. The appropriate and effective use of metadata in education agencies is one of those high interest areas.

The purpose of this guide is to empower people to more effectively use data as information. To accomplish this, the publication explains what metadata are; why metadata are critical to the development of 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 in an education organization. The primary audiences for this guide include technology, program, policymaking, administrative, and data staff in state and local education agencies. However, it may also be useful to other education stakeholders, including anyone engaged in operations or decisionmaking that depend on accurate, reliable, and timely information.

In This Guide

- Chapter 1 introduces the concept of metadata, or data about data, especially as related to education agencies and education data systems. Chapter 1 defines the term and explains why metadata are a critical component of sound education data systems and data management.
- Chapter 2 identifies common components of a metadata system for education organizations, including key points influencing metadata governance, metadata models, metadata item inventories, and comprehensive data dictionaries.
- Chapter 3 presents examples of metadata items commonly used by education organizations to improve data quality; technical operations; and data management, reporting, and use.
- Chapter 4 recommends planning processes and issues specific to metadata that contribute to the successful implementation of a metadata system in an education setting.
- Chapter 5 summarizes why it is imperative for education organizations to develop and implement a robust metadata system.
- Appendix A provides an example of standard definitions for common words that can be adapted to meet the needs of state and local education agencies around the nation. This type of tool helps to standardize data element names throughout an organization.



- Appendix B provides a real-world example of how a business rule gets developed.
- Appendix C defines the concept of a metadata registry and includes descriptions of several well-known metadata registries and standards available online.
- Appendix D provides a description of a metadata/business intelligence staff training program that can be adapted to meet the needs of state and local education agencies around the nation.
- Appendix E lists other resources related to metadata and education data quality, including sources referenced in the document and materials available from the National Forum on Education Statistics (the Forum), the National Center for Education Statistics (NCES), and other organizations.

The National Cooperative Education Statistics System

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 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. NCES provides management oversight of this work, 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 approves the document a final time, publications are subject to examination by members of the Forum standing committee sponsoring the project. The entire Forum (approximately 120 members) then reviews and formally votes to approve the document. NCES provides final review and approval prior to publication.

Table of Contents

Chapter 1

What Are Metadata and Why Are They Important?

This chapter introduces the concept of metadata, or data about data, especially as related to education agencies and education data systems. Chapter 1 defines the term and explains why metadata are a critical component of sound education data systems and data management.

Overview

We've all heard the warning about "comparing apples to oranges." But knowing that something is an "apple" or an "orange" is not always as simple as it seems. In the NASA Mars Orbiter accident, for example, engineers failed to recognize that the Orbiter's velocity system was measured in metric units ("apples") while its trajectory system was measured in English units ("oranges"). The

Exhibit 1.1

NASA's metric confusion caused Mars Orbiter loss

September 30, 1999

(CNN) – NASA lost a $125 million Mars orbiter because one engineering team used metric units while another used English units for a key spacecraft operation, according to a review finding released Thursday.

For that reason, information failed to transfer between the Mars Climate Orbiter spacecraft team at Lockheed Martin in Colorado and the mission navigation team in California. Lockheed Martin built the spacecraft.

"People sometimes make errors," said Edward Weiler, NASA's Associate Administrator for Space Science in a written statement.

"The problem here was not the error, it was the failure of NASA's systems engineering, and the checks and balances in our processes, to detect the error. That's why we lost the spacecraft."

The findings of an internal peer review panel at NASA's Jet Propulsion Laboratory (JPL) showed that the failed information transfer scrambled commands for maneuvering the spacecraft to place it in orbit around Mars. JPL oversaw the Climate Orbiter mission.

"Our inability to recognize and correct this simple error has had major implications," said JPL Director Edward Stone.

Source: *http://www.cnn.com/TECH/space/9909/30/mars.metric/index.html* results were truly catastrophic (see exhibit 1.1). While loss of life is not an issue with education data, serious problems may occur when data are used improperly, possibly affecting teachers' careers, school budgets, and most importantly, children's education.

Metadata[1] are defined as "data about data." A more technically precise definition is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage information." [2] In other words, metadata provide the context in which to interpret data and information. For example, in the case of the NASA Orbiter, metadata would have helped analysts determine the measures were metric and translate them into English units to avoid confusion and error. While the loss of the Orbiter is an extreme example of the value of metadata, countless data systems, including education data systems, could drastically improve data quality and data use by instituting robust metadata systems.

In the complex world of education data, answers to even apparently straightforward questions often depend on highly complicated and technical data. Take, for example, the "simple" question, *How many eighth grade English teachers are in your schools?* Exhibit 1.2 illustrates how the "correct" answer depends on the definition of each separate word or phrase in the question. On one end of the spectrum, there may not be any full-time certified English teachers teaching an English class to only eighth-grade students in the single middle school in the district this semester. At the same time, 50 or more full- or part-time teachers may be leading reading, writing, or language classes with at least one eighth-grade student at some point during the academic year. Clearly, the "right" answer depends on the context of the question and the data being used to answer it–and metadata provide that context.

Numerous cases in the field of education further illustrate the need for metadata. For example, consider a school superintendent's surprise when she stays up late preparing a presentation that uses real-time online data, then finds the results are different in front of her audience the next morning because the database was updated overnight. Access to metadata about the database's update cycle would surely have helped that superintendent. Or consider a state education agency that spent $1,000,000 on a new software application, but then cannot upgrade it without spending an additional $500,000 because the system was not properly documented during its original development. Metadata recording technical and management choices throughout the application's development would likely have solved this source code problem and allowed for less costly ugprades.

While using up-to-date data in a presentation and properly documenting technical specifications have always been important, the concept of metadata, or data about data, has never before been so relevant to educators. In this era of data-driven decisionmaking, education organizations and their constituencies place tremendous value on using data to inform instructional and management practices. In contrast, 20 years ago educators rarely used data for decisionmaking as a standard business practice; instead, school leaders often relied on impressions, opinions, and even instincts. Today's enhanced use of data is further complicated by the sheer volume of information collected. With more data to organize, access, and understand than ever before, a metadata system is an essential tool for accomplishing these vital information management tasks.

What Metadata Are Needed by Your Organization?

Many people in the school business need information to do their jobs. For example, a school principal needs to know how many students are in the ninth grade; a

Exhibit 1.2

Question: How many eighth grade English teachers are in your schools?

Component	Issues to be clarified by metadata
How many	Does "how many" refer to a head count or full-time equivalent (FTE) count?
Eighth grade	Does "eighth grade" include classes with seventh-, eighth-, and ninth-grade students; or just classes with only eighth graders?
English	Does "English" include reading and writing classes? Special education English language classes? Other language arts classes? English as a Second Language classes?
Teachers	Do "teachers" include only certified teachers? Only certified English teachers? Certified teaching assistants? Only teachers assigned to teach classes/students this grading period?
Are	At what point in time should the answer be valid? At the beginning or end of the current or previous school year?
In	Does this include teachers of students cross-enrolled in virtual settings? What if someone teaches English in more than one school—does he or she get counted more than once?
Your	Does this mean only schools under the authority of the state or local education agency, or does it include all schools within the boundaries of the state or locality?
Schools	Are special education schools included? Correctional institutions that grant educational degrees? Other residential facilities? Cross-enrolled virtual settings? Private schools?

state testing coordinator needs to know how well those students performed on an assessment; a local curriculum coordinator needs to know which of those students are taking advanced courses; and a superintendent needs to know how much it costs to educate those ninth graders. Accessing and interpreting these data requires a host of information management and technology metadata. Technical staff need to know where each piece of data is physically stored, and in what format. Other staff members, including program staff and the data steward, need to understand who owns each data set in the organization; as well as when the data were collected, what time period they represent, why they were collected, and how they are defined.

This type of information (i.e., the who, what, where, when, why, and how) is fundamental to the most basic operation of a data system; but many organizations, both within and outside of education, are unable to answer such basic questions about the data they maintain. The vast majority of organizations also cannot address some of the deeper, and in many cases more important, characteristics of data such as: Is the information private or otherwise sensitive? How are the data being used, if at all? Under what conditions are they valid for policymaking and reporting? How will pending changes in legislation affect current items, definitions, and code lists?

Exhibit 1.3

Questions about the data life cycle that metadata can address.

1 Needs Assessment
- What do you want to accomplish with data?
- What questions do you need to answer?
- How quickly do you need the data?
- In what format do the data need to be?

Collection
- Where did I get the data?
- When did I get the data?
- How did I get the data?
- Who owns the data?
- How are the data defined, derived, etc.?
- How do I know the data are valid?
- Why do I have the data (e.g., mandates)?

Storage/Security
- Where are the data?
- How can I find the data?
- In what format are the data?
- How did the data get there?
- Have the data been changed?
- Are the data private or otherwise sensitive?
- Does access need to be limited?

Analysis
- What do the data mean?
- How good is the data's quality?
- What are the data's limitations?
- How timely are the data?
- How can I use the data appropriately?
- How do others use the data?
- How do the data relate to other data?

Reporting
- What do the data mean to the reader?
- What was reported in the past?
- Why have the data changed?
- What may or may not be reported at the individual level? In aggregate?
- Which business rules govern report generation and data protection?

Retirement/Destruction
- Where will the data be archived?
- When do the data become invalid?
- What are the implications of preserving the data after that date?
- What procedures are required to destroy the data properly (e.g., "delete," shredding, degaussing)?

A wide range of working definitions exists for the term "metadata." Not everyone talking about metadata is referring to the same thing.

An example of a typical data and information life cycle is found in exhibit 1.3. When a piece of information is needed, it can be generated, usually as a result of a data collection or through derivation or other processes. The information then exists in storage or in use until, finally, it is retired, archived, or even destroyed depending on its sensitivity and ongoing validity (for example, certain health information, disciplinary records, and assessment scores may be destroyed after a student has left school). Metadata can describe the information at each stage in the cycle; in other words, a comprehensive metadata system can track a single piece of data (or a data set) as it evolves over time. These types of life cycle considerations have driven the development of metadata systems because the individuals who collect, maintain, and use the data require this information to effectively and efficiently manage data throughout its life cycle.

Perspective Influences "Metadata" Definitions

Most database managers, data stewards, education program managers, and librarians are probably familiar with the notion of metadata. However, given the different ways they view data—as something to be stored (the database manager), something to be maintained (the data steward), something to be used and reported (the program manager), or something to be catalogued and searched (the librarian)–it is hardly surprising that multiple definitions have arisen for the term "metadata" (see exhibit 1.4). These variations are not inconsequential because they complicate communication between staff, not only within a single education organization, but also across the education data community.

Metadata can also be viewed as an information management tool that transcends individual perspectives and, therefore, warrants a more comprehensive definition.

This comprehensive definition of metadata encompasses technical, management, retrieval, and usage perspectives; and serves as the working definition of metadata for the purpose of this publication. [3]

Metadata are structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

Exhibit 1.4

What are metadata?

- To a technology professional, metadata are "information about data warehouses (and other technical systems), including queries, reports, transformations, tables, columns, and users." (Adapted from Lee and Kim, A Metadata Oriented Architecture for Building a Datawarehouse, *Journal of Database Management.*)
- To a data professional, metadata are "Data about data, and not data in and of themselves; they describe the content, quality, condition, and other characteristics of the data." (Adapted from El-Sherbini, and Klim, Metadata and Cataloging Practices, *The Electronic Library.*)
- To a program manager, metadata are "All information necessary and sufficient to enable long-term secondary use (reuse) of a data set by the original collector, as well as by other users not directly involved in the original collection efforts." (Adapted from Michener, et al., Nongeospatial Metadata for the Ecological Sciences, *Ecological Applications.*)
- To a librarian, metadata are "Elements through which resources can be described and searched." (Adapted from Clyde, Metadata, *Teacher Librarian.*)

Complete references are listed in appendix E.

Metadata as a Component of Data Management

Metadata promise too much value as a business management tool to dismiss their implementation and maintenance effort.[4]

Metadata systems may not have been necessary when data sets were relatively small and simply organized. Under these circumstances, data were usually used by only a handful of people who were intimately familiar with each data element's definition, collection source, uses, limitations, and technical characteristics. Moreover, the

Some organizations rely on the experience of their data steward(s) as the primary source of information about their data. A metadata system is a better and more reliable alternative—and the only realistic way to effectively accomplish this vital information management task.

metadata that did exist often were stored in a data steward's memory or a program manager's paper files, and could be easily passed along from one person to another as a part of the organization's oral and written history. But the education enterprise has grown in complexity over the past decades, resulting in the seemingly exponential growth of information collected, stored, managed, used, and reported. In the field of education, as with other industries, metadata have become a necessary component of sound data systems. Without a formal and systematic method for conveying these "data about data," how can data, technical, and program staff confirm that information needed to understand the data will be available in a timely manner and appropriate format?

Metadata provide context for a single data item; serve as the backbone for efficient data management; and improve the use, analysis, and management of a body of data.

A well-managed metadata system minimizes disruption to data management and use. It ensures that the descriptions, definitions, parameters, usage instructions, and history of each element are maintained in an accurate and up-to-date manner. Additionally, metadata are essential for bridging programs and databases because they provide the framework for data exchange and communication within and between organizations. Metadata also inform data policymaking (for example, data retention procedures) and technology planning (such as load time demands) throughout an organization.

The benefits of properly implementing a robust metadata system include

- ✓ improving the likelihood of data meeting the users' information needs;
- ✓ improving the efficiency of data access and integration;
- ✓ improving the probability of correct data interpretation and use;
- ✓ identifying what data exist (and where) throughout an organization;
- ✓ identifying redundancy and disparity in data sets;
- ✓ increasing the efficiency of data storage and maintenance;
- ✓ improving the accuracy of data transfer across systems;
- ✓ improving the application of business rules and edit checks;
- ✓ reducing user expertise required to conduct effective queries;
- ✓ advancing data quality;
- ✓ ensuring the proper maintenance of information over time; and
- ✓ improving the quality of data-driven decisionmaking in the organization.

Despite its potential value, many organizations have not yet chosen to develop a thorough metadata system. Organizational leaders may make this decision passively if they are unaware of the need, or they may actively decide not to address this issue. Organizations that make an intentional decision not to develop a metadata system often do so because it would:

- ✓ demand expertise that staff may not possess;
- ✓ involve a great deal of work;
- ✓ take a lot of time;
- ✓ cost a fair amount of money;
- ✓ require a thorough understanding of current data resources;
- ✓ potentially expose existing deficiencies in data quality; and
- ✓ involve long-term commitment that does not match short-term goals.

All of these reasons for not developing metadata systems are valid–up to a point. Developing a system is a substantial undertaking that requires significant time, expertise, commitment, and money. But like other time-, staff-, and resource-intensive initiatives, such as installing new networking systems, or introducing new professional development programs, metadata systems should yield benefits that far outweigh the costs of implementation.

The consequences of neglecting metadata are many and severe. In the absence of a sound metadata system, the following types of serious data problems can, and often do, arise:

- ✓ a single data element may be applied inconsistently within an organization—for example, some staff members may code an absence reason as "excused" while others code the same reason as "unexcused";
- ✓ multiple conflicting definitions, code sets, and calculations may be used as though they are interchangeable even when they are not, such as different withdrawal codes or competing dropout rate formulas;
- ✓ a data value may be reported differently on different surveys—for example, different graduation rates may be reported for the same school because of different calculation dates or formulas;
- ✓ trend studies may not account for changes in definitions or policies that would otherwise influence analysis, such as changes in race/ethnicity categorization that might affect trends in student achievement;
- ✓ a data item, or even an entire collection, may be maintained when it no longer provides useful information, placing an unnecessary burden on data collectors;
- ✓ a new database may introduce terminology, definitions, and specifications that are not consistent with existing standards and protocols—for example, database designers may develop codes that will not be recognized by other users or systems in the organization;
- ✓ a data initiative may be at greater risk of failure due to unidentified data quality issues—for example, the implementation of a new data warehouse project may be inefficient because the underlying data quality is poor or insufficiently understood;
- ✓ policymakers may not thoroughly understand the data they are using—for example, they may not appreciate that there is a difference between the number of teachers expressed as a "head count" versus a "full-time equivalent" count; and
- ✓ data may be misinterpreted—for example, a graph of assessment results may seemingly show that student performance is improving when the apparent change is actually related to a new testing instrument.

In the past, some organizations have learned to live with these types of consequences. However, with the ever-increasing reliance on data for managing strategic and day-to-day decisionmaking, accepting these problems rarely is acceptable by today's organizational management standards. While metadata cannot eliminate every opportunity for incorrectly collecting, using, or reporting information, a sound metadata system provides a framework for better understanding data and, therefore, minimizes the likelihood of misuse. Exhibit 1.5 presents an example of the perils of data misuse and misreporting in an education organization.

You can't argue with the data

A false sense of security may arise when data are used improperly. Decisions based on misunderstood data can be disastrous. In fact, data without metadata can have consequences far worse than having no data at all.

Who Benefits From Metadata?

Although anyone who handles data or uses data for decisionmaking will benefit from the implementation of a sound metadata system, several categories of users have the most to gain.

For policymaking and administrative staff, metadata will help:

- ✓ improve data analysis and use by providing access to instructions and interpretation guidance;

Exhibit 15

Metadata: Solving the Case of the Inaccurate Dropout Count, Chapter 1

Everyone agreed that the new principal at Lincoln High School, Mr. Howell, was doing an excellent job. The teaching staff were well supported, students were motivated, and dropout prevention resources were properly targeted. The superintendent was therefore shocked to see that the school's dropout rate had increased dramatically over the past year. But how could the point be argued? The student population hadn't changed, the collection date hadn't changed, and counts at peer schools remained the same. Superintendent Sanders stared at the report in front of her.

Still enrolled:	4 percent
Completed:	35 percent
Dropout rate:	61 percent

How was she to explain what was happening at Lincoln High? A thorough review of the data was surely in order, so Dr. Sanders called Mr. Howell. The principal assured her that fewer students were dropping out at Lincoln High. "How can you be so sure?" Superintendent Sanders queried. "I can just tell," said Principal Howell. "I can't explain the report, but I know that our kids are staying in school." "Well, Mr. Howell," the superintendent replied, "I am sure you know what is going on in your school, but we can't just rely on your instincts. We have to look at the objective data, and they say that you're wrong… Now, let's dig into those numbers and see what is really going on."

This is the first installment of an ongoing story presented throughout this publication. *The saga continues on page 16.*

- ✓ reduce questions from the media and other data users by improving access to supporting information about data that are reported publicly;
- ✓ improve the accessibility and presentation of data for informing instructional and administrative decisionmaking;
- ✓ improve the likelihood that data about schools reflect what is really going on;
- ✓ identify why individual data elements are collected (e.g., information about mandates and use); and
- ✓ improve the understanding of connections between data and policymaking.

For technology staff, metadata will help:

- ✓ provide a clear list of technical attributes (for example, field length) that can be applied without having to reconsider management parameters each time a new item is collected and stored;
- ✓ improve the understanding of the business processes driving the collection and use of data that technical staff maintain;
- ✓ identify sensitive and confidential data, thereby improving system security;
- ✓ simplify and expedite data access and retrieval;
- ✓ reduce user inquiries to technical staff through improved system navigation and data accessibility; and
- ✓ simplify the exchange of data between systems, both within and outside the organization.

For program staff, metadata will help:

- ✓ reduce the likelihood of incorrect or inconsistent reporting;
- ✓ reduce collection demands by identifying redundant data elements;
- ✓ minimize questions from technical staff about data maintenance instructions;
- ✓ reduce questions from policymaking staff about data use instructions;
- ✓ improve data comparability and continuity over time within a program area and across the organization; and
- ✓ improve data auditing, thereby increasing overall data quality.

The Key Points of this Chapter…

- Metadata systems may not have been necessary when data sets were relatively small, simply organized, and used by only a handful of people. However, the education enterprise has grown in complexity over the past decades, resulting in the exponential growth of information collected, stored, managed, used, and reported by education organizations.
- Metadata provide the context needed to interpret data, and are most simply defined as "data about data." A more technically precise definition is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage information."
- High quality metadata deliver too much value to dismiss them as anything less than an essential business management tool. They are the backbone of efficient, accurate, and useful data systems.
- A robust metadata system improves the accuracy of data use and interpretation, as well as the efficiency of data access, transfer, and storage. The benefits are numerous for a wide range of education stakeholders, including policymakers, administrators, technologists, program staff, and instructional staff.
- Developing a metadata system is a substantial undertaking that requires significant time, expertise, commitment, and money; but a sound metadata system should yield benefits that far outweigh the costs of implementation.

Why should instructional staff, such as teachers and principals, care about metadata?

Data are frequently used to evaluate school, student, and even teacher performance. Anyone with an interest in accurate and fair school evaluation should have a corresponding interest in improving the accuracy and transparency of data collection, maintenance, reporting, and use—all of which can be derived from a robust metadata system. Furthermore, instructional staff and students benefit when metadata help make information needed to guide instruction available in a timely manner and useful format.

Notes:

[1] The term "metadata" was coined in 1969 by Jack E. Myers and trademarked in 1986 by his company, The Metadata Company *(http://www.metadata.com).* The trademarked version is written with a capital "M" and is distinguishable from public use of the term as "metadata" and "meta-data."

[2] As defined by the National Information Standards Organization (NISO), a nonprofit association accredited by the American National Standards Institute (ANSI) to identify, develop, maintain, and publish technical standards. *http://www.niso.org/* (Retrieved January 8, 2007.)

[3] Data consultant Michael Brackett avoids confusion over the term "metadata" by referring to the concept as "data resource data" and defining it as "any data necessary to thoroughly understand, formally manage, and fully utilize the data resource to meet the business information demand."

[4] Shankaranarayanan, G. and Even, A. (2006) The Metadata Enigma, *Communications of the ACM*, 49(2), 88-94.

Chapter 2

Description of a Metadata System

This chapter identifies common components of a metadata system for education organizations, including key points influencing metadata governance, metadata models, metadata item inventories, and comprehensive data dictionaries.

Exhibit 2.1

We're really *not* "just passing" those kids!

Most of the community grew up with a simple grading standard: you had to score at least a 70 percent to pass an exam. For years, the state assessment reflected the same standard, but a few years ago the content and scoring of the state exam were reconsidered. The assessment was made more challenging and the new cutoff for passing was set at 60 percent.

Unfortunately, when the test scores for the local schools made the newspaper, no mention was made of the redesigned and recalibrated state exam. The outcry was predictable: "No wonder we're making 'progress' in the schools… you're passing those kids with 60 percent marks… When I was a kid that would have never passed."

Although the school superintendent tried to explain that the exam had been made more rigorous, resulting in the new cutoff for passing, the damage had already been done… and no amount of public relations was going to correct it.

Metadata systems are driven by the information needs and characteristics of each specific organization (see exhibit 2.1). This makes it challenging, if not impossible, to describe a single model for a metadata system that would apply to every state and local education agency. Still, most metadata systems in education environments have some common features. The following description is based on these commonalities. In general terms, a robust metadata system will include

- system governance arrangements that include policies and procedures for metadata management and use within the organization, and related roles and responsibilities for staff;
- a metadata model that links metadata items to existing data elements and data sets;

- a list of relevant metadata items (i.e., a metadata item inventory), including a lexicon that identifies shared vocabulary for term use and naming conventions; and
- a comprehensive data dictionary.

Metadata System Governance

If a metadata system is to reflect an education agency's long-range vision, goals, and information needs, support for system development, use, and maintenance must exist at the highest levels of the organization. Management must also make sure that the organization's broader plans are considered adequately and that the metadata policy conforms to existing rules, regulations, and laws to which the organization is subject. And, of course, that adequate funding is budgeted.

Members of the organization's data governance team should consider metadata management to be as important as any other aspect of the organization's data system. As such, data ownership and stewardship responsibilities extend to metadata as well. Organizational leaders have a responsibility to ensure that all roles and duties for managing and using a metadata system are clearly delineated, assigned, and accepted throughout the organization. In addition to ensuring that staff members take their assigned responsibilities seriously, senior management should also develop and enforce policies and procedures that sustain the system and its use.

Many of the principles required for strong data and metadata governance can be found in *The Need for Data Governance*, available online at *http://www.ccsso.org/content/pdfs/EIMAC%20Brief%204%200308.pdf*.

Communication and accountability are as critical to metadata governance as they are to most operations in a school, district, or state education agency. Many organizations face a communications barrier between technical staff and data staff, often due to territorial, political, or data system evolution issues. When different programs, offices, or divisions within the same organization maintain separate data dictionaries, for example, the exchange of data across these entities will be limited or impossible. This communications barrier can be corrected by using universal data governance policies that require coordination, consistency, and standard protocols such as a unified data dictionary.

Metadata Managed Through a Metadata Model

A metadata model is a formal description of how metadata are structured to support the information needs of an organization. Like any data model, a metadata model can be described at a conceptual level, illustrating relationships between metadata items and the larger body of data around which they are generated; a logical level, reflecting the technical and operational parameters in which the metadata items exist; or from a physical perspective, specifying layout, file structures, and other characteristics. In more general terms, a metadata model is a representation of how an organization's metadata items relate to one another and to the larger body of data around which they are collected and generated. At a more detailed level, a metadata model maps and illustrates how data elements, metadata items, business rules, subsystems, data repositories, data flows, and information needs relate to one another. The type of system architecture used in the organization (see chapter 4) will strongly influence the appearance and logic of a metadata model.

Metadata Item Inventory

Most organizations with metadata systems maintain an inventory of metadata items. This inventory may be a spreadsheet developed by a data manager or a database created by the technology director. While the list of potential metadata items is quite long, most state and local education agencies focus on a subset that addresses *most* issues for *most* users. Exhibits 2.2a, 2.2b, and 2.2c present items likely to appear in such a list, organized into technical, data management, and usage/operations categories to illustrate common, though not mutually exclusive, usage.

Exhibit 2.2a

Common technical metadata items

Common name	Description
Field length	The recommended maximum number of places that the value of a data element would require in an electronic record system. For example, a descriptive alpha/numeric (AN) element might require 60 letters or numbers for a response, whereas a date (DT) would require 8 digits (MMDDYYYY). When designing a data collection system, both minimum and maximum lengths generally are specified.
Element type	A description of the form or qualities (i.e., the "type") of the data that constitutes the element. The *NCES Handbooks Online (http://nces.ed.gov/programs/handbook)* support several data element types, including alpha/numerical (AN), dates (DT), floating decimals (R), identifiers (ID), and numerical (N).
Permitted values	The range of possible acceptable values for a data field. For example, an elementary school may limit the permitted values for the Birthdate data element to reflect only the allowable age of elementary school students.
Code set	A list of choices that serves as a response for a data element. For example, "Female" and "Male" are options under the data element "Sex."
Translations	The transformation of a data value from one format, language, or presentation to another. For example, a date originally collected as 050806 (August 5, 2006 in the DDMMYY format) might be translated to 08052006 in the MMDDYYYY format in the target or destination system.
Storage/archival destination	The location (physical and/or electronic) where a piece of data is stored. This includes backup storage and should, as appropriate, be as specific as possible (e.g., the Blue Ridge Backup Facility, eastern wing, section 8, box 4, tape 2).
Data source	The collection instrument, data file, or formula from which data originated.
Data target (where used)	Any reporting instrument (including reports, report cards, publications, and other products), data file, or formula that uses or publish the data.

Exhibit 2.2b

Common data management metadata items

Common name	Description
Element name	The unique word or set of words that identifies the name of a metadata item (see appendix A).
Definition	A description of the meaning of a data element.
Purpose/mandate	The reason a data item is collected (e.g., state law, school board requirement, component of a report card indicator formula).
Restrictions	Any factors that limit the value, use, or interpretation of a data element. For example, data about a student's health conditions are often considered to be confidential and require appropriate access precautions.
Related data elements/components	Other data elements (or indicators) commonly used with the data element to enhance understanding or provide additional information. For example, all components needed to calculate a data element are included as related elements.
Calculations/formulas	The actual mathematical formula for computing a value.
Manipulation rules	A detailed description of the methods, techniques, and procedures for editing, deriving, calculating, translating, or otherwise manipulating a piece of data.
Ownership/stewardship	The individual or office that authorizes collection of the data and maintains responsibility for the attributes of a data element. Only this individual or office has the authority to change an attribute for that element and all subsequent use of that element should reflect authorized modifications.
Effective dates	The date a data element is introduced or modified, and the date its use ends in favor of a modification or retirement. All past start/end dates are retained as a part of a data element history.
Retention period	The amount of time a piece of data should be retained in active or archived form. A "disposal date" may be appropriate for data that will be destroyed.
Business rule	A rule under which an organization operates, and the expression of that rule as a mathematical or logical assertion governing how data can be entered or used within a data system. For example, a business rule in an education data system may say that values for the data element Age of Student must fall within the range of 5 through 21 (i.e., $5 \le$ Age of Student ≤ 21) if the agency serves only students who are between 5 and 21 years old.
Security/confidentiality	The classification for a piece of data that conveys the level of access and security to be applied to that data. In addition to the use of standard passwords, encryption techniques, and user authentication methods, security requirements sometimes specify the appropriate disposal of data. For example, a list of staff social security numbers may not simply be thrown in a trashcan or deleted from electronic disk storage but, instead, might require shredding of paper files or random binary overwriting for electronic files.

Exhibit 2.2c

Common data reporting/use metadata items

Common name	Description
Routine use	A description of the most common ways a data item is used appropriately. Conversely, this metadata item may also warn users about common ways the data are misused.
Key words	Any terms or phrases that relate to, or are cross-referenced with, an item (for example, related items for search functions).
Quality metrics	Measures intended to provide information regarding the relative quality of a piece or set of data. Examples of quality metrics might include completeness, continuity, contiguity, currency, reliability, accuracy, and coherence of a data set (see chapter 3).

The design of a metadata system is driven by the way an organization uses information. Thus, considerable variation may exist in different organizations' metadata item inventories, even within the field of education. The planning process discussed in chapter 4 places substantial emphasis on completing a needs assessment that gathers information from stakeholders about the data-related activities required for their jobs. A metadata item inventory should reflect and, in fact, be customized to meet those needs.

While reviewing available item inventories from peer organizations is advisable, system planners should not expect to meet their stakeholders' needs simply by copying another organization's item inventory (or the sample above) without any customization.

> While metadata offer many benefits for managing information, planners should resist the temptation to include too much information. Data about the metadata (meta-metadata) and even data about those data (meta-meta-metadata) may occur without vigilant management. Care must be taken to ensure that metadata items are limited only to stakeholder needs.

Data Dictionaries: A Critical Tool for Data Management

> If all the dictionaries in the world suddenly vanished, people would create as many different spellings and meanings for words as they could dream up—and disorder would overwhelm communications.

A data dictionary is an agreed-upon set of clearly and consistently defined elements, definitions, and attributes–and is indispensable to any information system. In the same way that standard English dictionaries help us use the English language effectively, data dictionaries help organizations maintain consistency in their information systems. Beyond data collection, database users and managers refer to a data dictionary to find out where specific data are located, whether they were reported correctly, how to use them appropriately, and what their values mean. Like an owner's manual, a data dictionary helps the data user understand and work with data.

Although many items in a data dictionary can be classified as metadata, data dictionaries and metadata systems are not identical. Data dictionaries generally contain only some of the metadata necessary for understanding and navigating data elements and databases and, thus, contain only a subset of the metadata found in a robust metadata system. Metadata systems, on the other hand, generally include the entire range of items used for data system management and analysis, including features for sorting, searching, organizing, and connecting data and metadata (see exhibit 2.3).

Exhibit 2 3

Metadata: Solving the Case of the Inaccurate Dropout Count, Chapter 2

Continued from page 8.

The school superintendent, Dr. Sanders, and Lincoln High School's principal, Mr. Howell, met on Thursday after school to study the dropout data. Despite Mr. Howell's assurance that more of his students were staying enrolled in school, the numbers hadn't changed:

Still enrolled:	4 percent
Completed:	35 percent
Dropout rate:	61 percent

Dr. Sanders quickly realized that they were in over their heads. "I can read the numbers in the report as well as anyone, but I have no way of knowing whether they are right. What report did you send to central office for us to generate this dropout rate?" Mr. Howell was embarrassed to admit he didn't know the answer off the top of his head. "To be honest with you, I have no idea… My administrative assistant, Ms. Johnson, fills out the reports." Dr. Sanders became irritated—not at Mr. Howell, but at the way the district was handling its data reporting. "You know, Mr. Howell, these data are very important. We use them to generate our attendance counts, dropout and graduation rates, and state funding estimates. And yet, you and I don't know how the numbers got on that form. We've got to change the way we handle data around here. In the meantime, let's get our data guru, Mr. Olsen, on the job."

A few minutes later, Mr. Olsen was in the superintendent's office, explaining that the data would have been entered on the year-end enrollment reports submitted by each school in the district. Dr. Sanders and Mr. Howell nodded their heads, as it all sounded vaguely familiar. Dr. Sanders asked the obvious question, "Is there any chance that we calculated the rates incorrectly, Mr. Olsen?" "Well, yes, that's always possible," Mr. Olsen responded. "But it's not likely. We verify our processes quite thoroughly. Without knowing more, I think that the numbers are either correct or the error is a result of something else." "Well," Mr. Howell interjected, "I don't think the numbers are correct, so we need to figure out what the 'something else' might be." "That's right," Dr. Sanders added. "We need to determine what is causing the error, and how we can be sure that it isn't happening in other aspects of our data reporting."

Continued on page 23.

The Key Points of this Chapter…

- Support for the development, use, and maintenance of a metadata system must exist at the highest levels of the education agency if the system is to reflect the organization's long-range vision, goals, and information needs.
- The organization's data governance team should consider metadata management as important as any other aspect of the organization's data system.
- A metadata model can portray a high-level perspective on the relationships between metadata and data, or it can depict a more detailed view of these relationships.
- Most organizations with metadata systems maintain an inventory of metadata items that address most issues for most users.
- A data dictionary is an agreed-upon set of clearly and consistently defined elements, definitions, and attributes–and is indispensable to any metadata system. Although many items in a data dictionary can be classified as metadata, data dictionaries differ from metadata systems in that they generally contain only a subset of items and features necessary to understand and navigate data elements and databases.

Chapter 3

This chapter presents examples of metadata items commonly used by education organizations to improve data quality; technical operations; and data management, reporting, and use.

Exhibit 3.1

Now these are useful metadata!

Debbie at the district office hated filling out the year-end financial reports for the state education agency. Information always seemed to go one way—to the state—without being useful to her district. Her outlook changed dramatically, however, when metadata in an error report arrived from the finance data manager at the state department of education. "Debbie," the note read, "our metadata system identifies any data that deviate from the state average by more than 12 percent. Your custodial costs were, on average, 18 percent higher than comparable districts across the state. Can you check into it and verify the costs you submitted?"

Debbie knew that her financial records were correct, but it didn't make sense that her district was paying 18 percent more for custodial services than comparable districts—her chief financial officer was far too shrewd for that! She agreed to review her submission, and quickly realized that she had used the wrong code set when querying the finance data system. The state had asked for a cost for supplies and salaries, but Debbie had given them the cost of supplies, salaries, *and* benefits. "Well, that would explain the difference," Debbie thought.

Unfortunately, Debbie had used the same number in her preliminary budgeting for the coming school year. "Wow, that correction will reduce the custodial costs in my budget! I am glad the state has a system to identify those types of mistakes… Maybe state reporting isn't a meaningless burden after all!"

A metadata system should help the organization in some tangible manner. Having metadata for its own sake (and not using it) is a poor use of resources. Therefore, an organization should identify its goals for metadata use. For example, metadata can be used to improve

- technical systems–e.g, quantifying the processing capacity of a search tool;
- data quality–e.g., ensuring a data element is defined the same way throughout the organization; and
- data reporting and use–e.g., verifying that all data values are within acceptable ranges.

Most organizations find many uses for metadata, especially in terms of improving the quality and use of their data (see exhibit 3.1). While this publication cannot list each and every possible application of metadata in an education setting, some common examples are listed in exhibit 3.2.

Exhibit 3.2

Commonly used categories of metadata items*

Technical	Data management	Data reporting/use
Field length	Meaning	Identity
Element type	Availability	Accuracy
Permitted values	Restrictions	Reliability
Code set	Limitation	Completeness
Translations	Components/operations	Sparsity
Storage/archival location	Purpose/rationale	Value set testing
Source	Owner	Coherence
Target	Steward	Continuity
Load time	Time parameters	Contiguity
	Treatment	Currency
	History	Punctuality
	Retention	Verification
	Security/Confidentiality	Validation

*See text and exhibit 3.3 for descriptions.

Technical Metadata

The most basic technical metadata items are collectively know as "data attributes," which are technical specifications and parameters that inform how a piece of data is designed within a technical system. Data attributes include a data element's **field length** (e.g., up to 12 characters), **element type** (alphanumeric, date, etc.), **permitted values** (such as, 0-999 inclusive), **code sets** (e.g., M = male and F = female), and technical **translations** (e.g., changing date data from a DDMMYY to a MMDDYYYY format). More information about data attributes commonly used in the field of education can be found in the *NCES Handbooks Online* at *http://nces.ed.gov/programs/handbook.*

Storage location identifies the physical and/or electronic location where data are stored. This includes a building site, such as "in office #213" or "at the offsite storage facility at 123 Jones Street"; the machine, such as "computer/server serial number 1234"; and the database, table, and column, such as, "staff_db, assignment_tbl." Because data do not just appear in a data system and stay there indefinitely, other useful sets of technical metadata are data source and data target. Data **source** refers to information that identifies where data came from–either technically (e.g., a particular database) or operationally (e.g., a particular survey). Data **target,** on the other hand, is a description of the data's predicted destination, such as another database or a report.

These metadata are critical when programmers are designing extract, transform, and load (ETL) processes that move data from one system to another.

For some types of datasets and processes, **load time** can be important metadata. When processing capabilities are strong or data loads are simple and range from milliseconds to one or two seconds, load time may not be worth measuring. But a school district that loads 200,000 attendance records each morning needs to know when the system is going to be engaged at full capacity for a couple of hours.

Data Management Metadata

At their most basic level, metadata are intended to convey information about data **meaning** Management items include a data element name, definition, and other data dictionary entries necessary for understanding the meaning and context of any single piece of data. For example, the Number of Graduates in many states includes only those students receiving regular, standard, endorsed, or advanced diplomas. In other states, however, the Number of Completers is used to count graduates as well as students who receive a high school equivalency certificate, certificate of completion, or attendance certificate. The relative meaning of these data, therefore, clearly depends on how the terms "graduate" and "completer" are defined, and anyone using the information would benefit from metadata that provide clear and accurate definitions for the terms.

Data users often find themselves concerned about data **availability,** which can be presented as a catalog of what and when data are available. Availability may vary for different users. For example, data might have an earlier release date for internal planning than for external public reporting.

Restrictions and **limitations** help users identify factors that limit the use, value, or interpretation of a data element. Restrictions might include privacy/sensitivity labels warning users not to share data, or indications about combinations of data that may not be released, such as name and assessment scores. Limitations often address more practical issues, such as a non-comparability warning about two apparently similar items that should not be compared because of meaningful differences in sampling techniques. More advanced users might be interested in data **components/operations** that describe how a data value was generated based on its components and derivations as in, for example, what data elements and what formula were used to generate a dropout rate. Data **purpose/rationale** generally indicates the underlying reason for collecting the data, including public laws or administrative policies that require collection.

One person or office in the organization should be responsible for defining each data element and assigning access rights to it. Many organizations call the person or office with these responsibilities the data **owner** A data **steward,** on the other hand, is the individual or office accountable for maintaining a data element's definition and metadata in a manner consistent with the rules established by the data owner. In other words, a data steward works on behalf of a data owner. While the labels "data owner" and "data steward" may vary across organizations depending on governance structures, management terminology, and organization size, the distinction between decisionmaking responsibilities (data ownership) and management responsibilities (data stewardship) is critical to the effective operation of a data and metadata system.

Data owners are responsible for determining domains that define the range of permitted values (e.g., 1–999 inclusive). They are also responsible fort the data's **time**

parameters–information about the date when the data were collected or loaded, and the period for which the data are valid.

Data **treatment** describes how data were modified or otherwise changed, in format or presentation, after collection. This includes information about mapping and transformations, as well as rules for significant digits, rounding, cell sizes, business rules, aggregating, and other formulas and derivations. Data **history** is often presented in the form of an audit trail or other record of how, when, and why data were modified, and by whom.

As an extension of data storage, **retention** metadata indicate how long data should be maintained, and when and how they should be destroyed at the end of their life cycle. For example, some enrollment and fiscal data are maintained indefinitely as a function of historical recordkeeping for a school, district, or state; however, private student information such as health and disciplinary records may need to be destroyed as soon as a student is no longer enrolled in school. **Security/confidentiality** metadata items are often used to identify sensitive and private data. If, for example, data such as social security numbers are identified as particularly sensitive, appropriate destruction methods might include sophisticated technologies such as degaussing (neutralizing the magnetic field of storage tapes) or binary code overwriting.

Data Reporting/Use Metadata

Quality is a complex, yet critical theme in data production and use. Individuals using data for organizational decisionmaking, program evaluation, or research must understand the quality of the information they rely upon. A host of related concepts, including a wide range of quality metrics, are often used as metadata for assessing and tracking the quality of a data element or data set. One measure that directly assesses a data set's quality is **identity,** which is used to determine whether every "item" (e.g., a person, place, concept, or event) is uniquely identifiable and distinguishable from all other entities in a data set. Identity analysis frequently addresses the following types of issues:

- The same identity key is shared by more than one entity; for example, the same Social Security Number is used by two or more staff members.
- The same entity has more than one identity key, such as a single student with more than one student identification number.
- Entities have missing or incomplete identity keys; for instance, a class without a course identification number.

Accuracy and **reliability** are also directly related to data quality. Accuracy metrics determine the extent to which data measure what they purport to measure without bias. In other words, how well do the data correspond to the process or product being assessed? For example, an accuracy metric could help determine whether an exam assesses academic performance without introducing bias. Reliability, on the other hand, refers to the consistency, reproducibility, and dependability of the data. If the same item were measured multiple times, would the same results be generated? Reliability may reflect uncertainty in a measurement tool or the amount of random error naturally present in the data.

Completeness measures the degree to which required records and values exist in a given data set. For example, if individual student records are being transferred, the record set is considered "complete" when a unique record exists for each student in the group; if there are 200 students, the record set is complete if there are 200 unique records. Similarly, if there are 50 mandatory items or fields in each individual student record, a record is complete when each of the 50 fields has an entry. Because completeness is determined by having an entry in each field, all data items must be completed unless a skip pattern (or similar tool) is used for items that need not be completed.

The inverse measure of completeness is the concept of **sparsity** This refers to a measure of a lack of data when, for example, only four of nine required fields are available. When data are too sparse, assessing what they mean becomes difficult.

Value set testing examines the content of data fields to ensure that each data value falls within the expressed domain of allowable values. Allowable values (e.g., the age of all students in an elementary grade level must be between the values of five and twelve) are often based on business rules and other guidelines and standards expressed in metadata. The frequency or rate of domain violations and percentage of defective values are the most common measures of value set integrity. **Coherence** complements value set testing by providing a measure of value conflicts across related data sets. In other words, not only do data fall within a range of allowable values (value set testing), but data that should be identical in different data sets are indeed the same. For example, are student counts on an annual enrollment collection consistent with student counts in an annual dropout report?

Another facet of data quality is **continuity** analysis, which typically is performed to confirm a consecutive, non-overlapping, and unbroken history of the events represented by the data. For example, continuity analysis might assess whether daily membership data are available for each school day (and, in fact, only once for each day) in an academic year prior to generating an average daily membership for the entire year. If average daily membership were to be generated for each grading period, these data would need to be available consecutively from the first through the last school day of the grading period. Common continuity measures include the ratio of entities with a defective history to those with a defect-free history. More complex measures examine the size of the gap or overlap when defects occur.

Contiguity testing further assesses the logical sequencing of data in a data set. For example, contiguity measures might be used to assess whether the date a student passes the state's exit exam always occurs prior to the date of graduation. Contiguity evaluation generally is based on business rules, as well as other guidelines and standards expressed in metadata, to define the logic against which data are assessed. Typical contiguity measures include the ratio of entities with a defective history to entities with a defect-free history. More complex measures examine the frequency with which particular steps in a required sequence are skipped or recorded out of order.

Currency refers to the age or "freshness" of the data—that is, how "current" it is. Currency usually represents the time difference between the present date and the date when data were entered in the database. It often is measured in terms of the gap (the number of hours, days, months, or years) between the current date and the date of the most recent data available. This type of information is most important when great changes in data values can occur over short periods of time, or when data are used routinely but not collected very frequently. The effect on the end user can be significant—for example, a user should know if the "latest" enrollment data were collected eight months previously.

An extension of currency is **punctuality**, which is a measure of how quickly access is provided to recent data. For example, if student addresses are updated in May, when

Exhibit 3.3

Examples of metadata for the data element student birthdate

Metadata item	Example of metadata maintained for Student Birthdate
Field length	8
Element type	DT (date)
Permitted values	Student Birthdate, when collected, should not be earlier than the date that would result in the student being over age 20 if that student is enrolling in grade levels preK to 12.
Translations	These data are available to authorized viewers in the operational data system, but are otherwise encrypted (via master algorithm) and suppressed in all public reporting.
Storage location	Server = svr10079prod; database = Student_Information; table = sat_student_core; field = birth_date
Source	Student Enrollment Collection System
Target	Provided for fiscal auditing and internal management purposes, and used in data verification audit processes.
Meaning	The month, day, and year an individual was born.
Restrictions	Only users with access to individual student data are permitted to view this element.
Limitations	This item does not necessarily reflect the student's grade level.
Components/operations	The element can be compared to the current date to calculate a student's age.
Purpose/rationale	To serve as the district's principal method for determining a student's age. Also used in matching criteria to identify student.
Owner	District Registrar
Steward	Element is managed by Enrollment Specialist (Mary Smith) and backup is Senior Business Analyst (James Brown)
Time parameters	Student Birthdate is active upon assignment at enrollment and continues until all individual records are removed from the system.
Treatment	Birthdates entered in alternative formats (MM/DD/YY or name of month, day, and two-digit year) are converted into a MM/DD/YYYY format.
History	Once entered, the element is not changed for an individual student.
Retention	Five years after student has exited the school district.
Security/confidentiality	Sensitive and confidential.
Identity	Each individual may only have one birthdate on record.
Accuracy	Audited once after original entry.
Completeness/sparsity	94 percent of the 2007–2008 records loaded contain values for this field.
Value set	89 percent of the 2007–2008 records loaded contain values within the domain of permitted values.

Note: These entries are presented as examples and do not represent metadata from an actual school, district, or state data system. Some types of metadata described in this chapter are more appropriate for describing sets of data rather than individual data elements; these are not included in this exhibit.

are they available to the transportation office for planning the following school year's bus routes? Punctuality is sometimes referred to as timeliness (are the data available for use when needed) and may also be used to establish schedules that describe when new data can be expected. The punctuality measure may vary for the same set of data depending on audience type; for example, a data set may be available for internal planning purposes more quickly than for external reporting.

Data **verification** is the practice of confirming that data are accurate, and data **validation** refers to the practice of confirming that data agree with expectations of reasonable values and accepted norms. These are integral concepts in the production of quality data (see exhibit 3.4). Metadata can document the results of various statistical and procedural techniques used to verify and validate data. These include response and documentation audits, such as an examination of records that substantiate data submitted by a respondent; cross-checks, which refers to the practice of "checking" data from different collections for consistency; and value edits, which, for example, can compare entered data to maximum or minimum expected values. Exhibit 3.3 provides a real world example of how these metadata concepts might be applied to a data element in a metadata system.

Data Profiling

A good place to end a discussion on quality metadata is with the concept of a data profile. A "profile" is commonly defined as "an analysis representing the extent to which something exhibits various characteristics" *(www.thefreedictionary.com/profile)*. As an extension of this idea, a "data profile" is a formal summary of distinctive features or characteristics of a data set, including the data quality items described throughout this section.

Data profiling generally starts with an examination of what an organization expects to find in its data (or database), and then determines whether the data reflect those expectations. For example, what percentage of fields contain data? If a field is mandatory, it should be 100 percent, but profiling may uncover a somewhat different reality. Similarly, if a field stores a coded value, what and how many codes are found

Exhibit 3 4

Metadata: Solving the Case of the Inaccurate Dropout Count, Chapter 3

Continued from page 16.

Mr. Olsen, the district's data guru, spent the next day reviewing each step in the process used to generate the state dropout report. He traced Lincoln High's data from the moment they entered his system to their submission to the state education agency. He checked the dropout calculation formula, looked for rounding errors, reran edit checks, examined verification procedures, and compared the results to those of peer schools in the district. After seven hours of diligent review, he was pleased to announce with confidence to Superintendent Sanders and Principal Howell that the error had not originated in his office. While they were pleased to hear that the district's data team was doing its job correctly, the two educators were disappointed that the problem hadn't been solved. "So if the error didn't originate at the district level," Principal Howell said dejectedly, "it came from my school." "I'm afraid that seems to be the case," Superintendent Sanders agreed. "Okay, then," Mr. Howell replied resolutely, "we'll go through our data submission process tomorrow from top to bottom to locate the problem."

Continued on page 31.

in that field? For example, in some organizations, sex might be represented within a single database by "F, M," "female, male," or "x, y." More advanced data profiling techniques can determine whether a particular information system tends to over- or undercount some feature in the dataset, (e.g., the number of students) relative to expected results. As such, profiling often is used to evaluate data quality; assess whether a collection system supports quality; and determine whether documentation and other available guidance are being used correctly.

Business Rules

Business rules have been defined as both "directive(s) intended to influence or guide business behavior" and "constraints on a business."[1] In other words, they describe what an organization *must* do or, alternatively, what it *cannot* do.[2] Business rules are a form of metadata and, in this capacity, express the guidelines an organization has established for using or modifying a particular data element or data set (see appendix B).

Virtually all organizations have rules, whether they are a fortune 500 company, a public elementary school, a local fast food restaurant, or a family of four. These rules, frequently referred to as an organization's "policies," can range from the informal "No Shirt, No Shoes, No Service" or "No dessert unless you finish your vegetables," to much more formal and specific guidelines such as, "All records of students in grades 3–11 must have a valid score on the annual state math assessment." Some organizations express their business rules in natural, informal language, while others choose to use more formal styles. Regardless of their tone, good business rules should meet certain criteria (see exhibit 3.5). They should

- be explicitly expressed, either in formal language or graphic representation;
- follow an adopted standard for expressing all business rules; and
- be declarative, describing a required or prohibited state.[3]

These declarations should be stand-alone and absolute statements of truth about how the organization operates. In other words, one should not be able to break them down into simpler statements and they should be interpretable under any circumstance as either completely true or completely false. For example, the business rule "student age cannot exceed 24 years as of September 1 of the current year" means that the age recorded for a student must, under all circumstances, be less than or equal to the value of 24 years old as of September 1; and any value in an age field is either completely consistent or completely inconsistent with this rule.

As with any skill, applying metadata to business operations begins with learning basic steps before progressing to more advanced functions. This introduction to metadata concepts illustrates the types of useful information that metadata can provide education organizations. It is not an exhaustive list or a detailed description of how to use these powerful information management tools.

Exhibit 3.5

A real-world education business rule

The Tennessee Department of Education's funding formula for its Basic Education Program (BEP) provides an example of a business rule in the field of education data. Consider the following information about how Average Daily Membership is structured in the Tennessee BEP:

Component ➡	**Average Daily Membership (ADM)**
Structural assertion ➡	A student enrolled or receiving instructional service is included in an ADM calculation with the exception of the following: ∘ Out-of-state, non-resident student ∘ I-20 student (student from out of the country attending a Tennessee public school) ∘ PreKindergarten
Action assertion ➡	No student may earn more than 1.0 ADM. The formula for ADM is designed to handle part-time and vocational students accurately. ADM for vocational students must be calculated separately for each vocational program in which the student is enrolled. Formula used to calculate ADM is:
Derivation ➡	ADM = Σ [Days Scheduled * Time Scheduled ÷ Standard Day] ÷ Report Period Days
Additional structural assertions ➡	The symbol "Σ" means "sum of all values for the equation between the [brackets]." If the student's time scheduled and standard day never change during the report period, the formula inside the brackets needs to be calculated only once. Days Scheduled is the number of days a student is enrolled with a particular Time-Scheduled during a report period. (See note below.) Time Scheduled is the lesser of, either the total number of minutes, excluding lunch, a student is scheduled for classes for the school day; or the Standard–Day. (See note below.) Standard Day is the length in minutes of a standard, full-time day of instruction, excluding lunch, for all students, regardless of whether they are enrolled full-time or part-time. (See note below.) Report Period Days is the number of days covered in the report. Standard report periods are 8 successive, 20 instructional-day periods; a ninth period, which may be 20 instructional days or less; and the full school year, which may be 180 instructional days or less. A full-time student is a student whose amount of time receiving instructional service is greater than, or equal to, the amount of time specified in the Standard Day. A part-time student is a student whose amount of time receiving instructional service is less than the amount of time specified in the Standard Day. NOTE: For full-time students, and for schools that do not have part-time students and do not schedule students into individual classes, such as elementary schools, Time Scheduled equals Standard Day.

The Key Points of this Chapter…

- Metadata are useful only when they help an organization in a tangible manner. From a practical perspective, metadata can be used to assess and improve technical systems, data management, and data reporting and use.
- Data profiling uses many of the metadata items in this chapter to characterize the quality and other features of a data set.
- Business rules describe the policies and guidelines an organization has established to manage its businesses processes, data elements, and data sets related to those processes.

Notes

[1] Ross, R.G., *Principles of the Business Rules Approach*. Addison Wesley Professional, 2003.

[2] Hay, D.C., A Repository Model – Business Rules, Part II (action assertions). Available at: *http://www.tdan.com/i020ht02.htm*.

[3] Perkins, A., Business Rules = Meta-Data. *Technology of Object-Oriented Languages and Systems, TOOLS 34*, 2000.

Chapter 4

Implementing a Metadata System

This chapter recommends planning processes and issues specific to metadata that contribute to the successful implementation of a metadata system in an education setting.

Exhibit 4.1

Metadata 101: Metadata don't fix broken data systems

Karen Hall, the district's data steward had high expectations for the new metadata system. But the astute data expert noticed something strange happening during installation, and she began to worry.

As Karen helped the vendor map the district's data elements and data sets to the new system, she identified numerous mistakes in her data's format, structure, and logic. Invariably, the vendor had the same response whenever a problem presented itself: "You don't want us to deal with that… It would take way too much time to straighten it out and would delay the implementation of the new system."

The first couple of times this happened, Karen kept a mental note of what needed to be corrected in the system, assuming she would fix the problems at a later date. But by the time the list had grown too long to trust her memory, she decided to raise the issue more formally with the vendor.

She was very direct: "So how will this metadata system work when it can't be mapped to our data in a consistent way?" After observing the vendor's evasive reply, Karen went straight to her point, "How can a metadata system work properly when the main data system isn't configured consistently?" After doing some research on her own, the data steward learned what she had suspected all along: If you don't have a clear sense of the data in your system, you can't expect a metadata system to help you better use and manage the data.

Karen then gave the vendor a choice: take the system's data quality issues seriously, or have the contract canceled so the RFP could be revised to include system analysis prior to metadata implementation.

Introducing a metadata system is a complex endeavor that requires planning comparable to any other large organizational initiative (see exhibit 4.1). The concepts outlined in this chapter focus on the steps that are particularly critical for, or unique to, planning and implementing a metadata system in an education setting.

Metadata systems are built around existing data systems and, ideally, the organization's vision for future data use and management. As such, their development should be driven by the information and business needs of the organization. In other words, what do data users need to know to effectively manage and maximize the quality and utility of their work? A thorough planning process that incorporates data, technical, administrative, and management perspectives improves the likelihood that the system will meet user needs and organizational goals.

Establishing a Planning Team

If the organization's staff do not possess metadata expertise, the first step in building capacity will be to train prospective team members about metadata and the potential benefits of using metadata in education organizations.

Whether the system is developed from scratch or an off-the-shelf purchase, planning is a time-consuming task that requires considerable data and technical expertise, a thorough understanding of the organization and its data operations, and extensive project management skills. As such, a planning team should be established to set the course for the project. Team members will likely include the organization's data manager, a technical authority, and a representative from the organization's data governance body. In addition, they should also include representatives of other stakeholders who will eventually use the system, such as data entry staff, data analysts, program staff, and policymakers responsible for data-driven decisionmaking. The team should have executive sponsorship and be led by a project manager with proven leadership skills and sufficient authority to direct the team and make day-to-day decisions without additional permission.

Conducting a Metadata Needs Assessment

One of the challenges faced by the planning team is to implement a metadata system that meets the needs of many different types of users, collectively referred to as "stakeholders." Meeting their needs will require a comprehensive "needs assessment," which is undertaken to gather information about how stakeholders will use a metadata system so that planners can, in turn, ensure that the system being developed will meet those requirements. However, because many stakeholders may not be familiar with the concept and potential advantages of metadata, beginning the needs assessment with examples of metadata use may be helpful so that stakeholders gain enough understanding to provide meaningful input.

Planners must be able to distinguish between "wants," those features that stakeholders would like to have; and "needs," those features that are required to run the organization.

The end product of a needs assessment is a Needs Statement. A good approach for documenting the Needs Statement is to write it as though all staff involved in its creation will be taken off the project and new people will implement the next phase. Documentation can be considered effective if the "new people" can read the Needs Statement as a stand-alone product and understand its findings without additional input from the "old" team.

A Needs Statement should describe both functional needs and technical needs (see exhibit 4.2). In this context, functional needs are defined as the tasks or actions, or "functions," the metadata system will accomplish. For example, functions may include

- ✓ locating the definitions and other attributes of all metadata items in the system;
- ✓ entering metadata into the system;
- ✓ searching by key words and terms;
- ✓ customizing and generating metadata reports;
- ✓ harmonizing with the data dictionary;
- ✓ linking to external data standards;

- ✓ updating metadata items;
- ✓ identifying metadata item modification history;
- ✓ mapping metadata items to individual data elements;
- ✓ identifying data element "owners" and "stewards";
- ✓ enabling data "owners" and "stewards" to modify data and metadata;
- ✓ mapping data items to their physical storage location within a data system;
- ✓ assessing data quality; and
- ✓ regulating system access.

The technical needs included in the needs statement should not be overly technical or complex. They are simply statements of capabilities required of the technology solution that will support the metadata system. Capabilities might include

- ✓ meeting all relevant technical standards and specifications;
- ✓ accomplishing expected performance requirements;
- ✓ achieving ease of use/interface expectations;
- ✓ providing access safeguards and security for sensitive and confidential information;
- ✓ handling peak user-capacity;
- ✓ accommodating connection needs for users based on their location and how often they need to access the system;
- ✓ maintaining version control for the data dictionary and business rules; and
- ✓ automating loading and updating capabilities.

Exhibit 4.2

Suggested outline for a metadata system needs statement

Section 1. Introduction
- 1.1 Background
- 1.2 Objectives and scope

Section 2. System contents
- 2.1 Types of metadata
  - 2.1.1 Data management
  - 2.1.2 Reporting and use
  - 2.1.3 Technical
- 2.2 Metadata standards
- 2.3 Volume of information

Section 3. System functions
- 3.1 Storage and retrieval capabilities
- 3.2 Calculation and processing capabilities
- 3.3 Collection and output capabilities

Section 4. Access and capacity
- 4.1 Interface requirements
  - 4.1.1 Web, WPN, PC/Mac, etc.
- 4.2 Hours of operation
- 4.3 Number of users
- 4.4 Transmission volume
- 4.5 Security and access requirements
  - 4.5.1 User categories
  - 4.5.2 Permission restrictions
  - 4.5.3 Remote access

Section 5. Technical parameters
- 5.1 Adherence to technical standards
- 5.2 Requirements for system interfaces

Exhibit 4.3

Clothes are commonly stored in two ways: neatly or haphazardly. Either way meets the ultimate need of "storing" clothes, but one system is more efficient and more useful. Choosing the efficient approach doesn't mean you have to go to the extreme and sort each article of clothing by color or age, but a closet that allows you to organize by seasonal utility might be beneficial. With respect to metadata, the amount of effort expended to organize data in your metadata system (like clothes in your closet) should be determined by what makes sense for you and your users.

Incorporating Relevant Metadata Standards

Adopting generally accepted standards into a metadata system has the potential to yield many benefits. An organization gains immediate access to expertise shared by the standards' publishers rather than waiting for staff to develop comparable levels of expertise through training or trial-and-error. Staff can quickly assimilate existing standards and the expertise they reflect although, admittedly, not always with the developer's depth of understanding. Similarly, incorporating existing standards decreases the time needed to develop a new system. Rather than starting from scratch, standards can help frame a development project and provide a template for a starting point. Finally, using accepted standards improves the comparability of an organization's data with external systems and partners within the field of elementary/secondary education; within a single governmental unit such as a state or county; or even with specific institutions that commonly exchange data with the organization, such as colleges and universities. See appendix C for more information about metadata standards and metadata registries.

Conducting Cost–Benefit Analysis and Estimating Return On Investment

Metadata can reduce reporting burden, make data more accessible, and improve data quality—which can all result in more time to teach and otherwise support students.

Deciding whether to proceed with the development of a metadata system will eventually be based on two questions:

- How much will the metadata system cost?
- Will the benefits outweigh the cost?

To answer these important and appropriate questions, planners engage in cost–benefit analysis to ensure that both the positive and negative implications of a metadata system have been considered. As an extension of cost–benefit analysis, "return-on-investment" (ROI) is a concept used to express the amount of benefit (return) relative to the amount of resources (investment costs) needed to produce the return. Based on thorough analyses, many organizations have found that the potential improvements to data quality and use are well worth the costs of developing and implementing a metadata system.

In addition to costs for hardware and software, staff and/or consulting, and other direct development requirements, planners should also expect indirect costs. These are often referred to as "unanticipated costs," although many of can them can be anticipated with careful planning. These types of costs include staff training (initial and ongoing), user support (help desks, tutorial development, etc.), system maintenance costs, licensing agreements, and ongoing system evaluation initiatives.

Exhibit 4 4

Metadata: Solving the Case of the Inaccurate Dropout Count, Chapter 4

Continued from page 23.

The next morning, Mr. Howell called a meeting of Lincoln High School's administrative staff. "Okay, we've got a data problem," he said. He saw a few faces glaze over and decided to emphasize how important this work was, "But don't you dare think this is an unimportant administrative chore." He showed his staff the faulty dropout data submitted to the state. His team members sat up, troubled by the numbers that so unfairly represented their hard work resolving Lincoln's dropout problem. "But that isn't our dropout rate," the vice principal interjected. "I know that and you know that," Principal Howell answered. "But that's what the rest of the state will think unless we figure out where we made a mistake… That's right, the problem originated here, and we need to get it straightened out."

Four hours later, the administrative leaders were still stumped. "I just don't get it," the vice principal said with a grimace. "The numbers add up correctly on the form, but somehow they result in a dropout rate that is incorrect. We've gone over this three times, and nothing is changing." At that very second, Ms. Johnson, the administrative assistant, began to stammer. "Hold on… hold on… hold on…" she said with growing conviction. "Just hold on and look at this." All eyes turned to her as she explained what she had found. "I went back to the original coding instructions. They're not the same codes we use." And sure enough, they weren't. Ms. Johnson continued. "Notice, the first two codes are the same… I bet someone saw that and just assumed the rest were the same as they have always been for the data we send the district."

Lincoln High School Exit Codes	Dropout Report Exit Codes
1. Still enrolled	1. Still enrolled
2. Transferred	2. Transferred
3. Completed	3. Dropped out
4. Dropped out	4. Completed
5. Exited–neither completed nor dropped out	5. Not enrolled, eligible to return
6. Other	6. Exited–neither completed nor dropped out

Principal Howell called Mr. Olsen, the district's data manager, and asked him to check the codes. Mr. Olsen realized immediately that this was probably the root of the problem. "Yeah, the state adopted a new code set. I'm sure we sent out a notice, but evidently it fell through the cracks somewhere along the way. I'll run the numbers with the changed codes, but I'd be very surprised if this doesn't fix things." Principal Howell and Mr. Olsen let out a sigh of relief and looked forward to sharing the good news with Superintendent Sanders.

Continued on page 43.

The absence of a market price for good data presents a challenge to cost–benefit analysis for metadata systems. However, some cost savings from improved data quality can be measured in the areas of purchasing, staff allocation, and maintenance and operations. Cost avoidance may also be factored into the analysis; this might include not needing to hire consultants or purchase products to revamp aspects of your data system.

Some of the benefits of a metadata system are easily quantifiable, but many are not (see our ongoing story, Solving the Case of the Inaccurate Dropout Count). Still, potential financial benefits can be estimated. We know, for example, that a robust metadata system can reduce redundancy in a data system; this, in turn, should decrease collection, access, and reporting burdens—each with a tangible cost. Similarly, metadata systems can make data more accessible, saving staff time. Metadata also improve data

Exhibit 4.5

Example of metadata system cost–benefit and return-on-investment (ROI) analysis

Costs		Benefits*	
Hardware and software	Purchase of the computers, networking equipment, and software needed to operate the system.	**Reduced IT costs**	Savings associated with reduced technical demands because of efficiencies resulting from a metadata system; for example, redundant data are removed, decreasing storage needs.
Installation	Payment to in-house staff or outside contractors to install the system.	**Interoperability**	Savings associated with improved effectiveness and efficiency when sharing data across two or more systems.
Consulting	Payment to outside contractors for technical or other expertise during system development, installation, implementation, and training.	**Productivity gains**	Savings associated with increased staff output and efficiency because of improved data access and understanding.
Initial training	Costs, including participant/staff time and logistical expenses, associated with providing introductory system training.	**Reduced data burden**	Savings related to a reduction in the resources (e.g., staff time, collection demands, and reporting effort) required to collect, manage, or report data.
Ongoing training	Costs, including participant/staff time and logistical expenses, associated with providing ongoing training.	**Reduced redundancy**	Savings associated with reducing unnecessary data (e.g., data that are no longer used). The "Rule of Ten" warns that one in ten data items are redundant in many large data systems.
Staffing changes	Costs associated with reassigning staff duties because of system maintenance or usage requirements.	**Data quality**	Savings associated with improving the validity, reliability, utility, and timeliness of data, such as decreased auditing costs.
Support and maintenance	Costs to maintain a system over time, such as upgrades, routine maintenance, and malfunctions.	**Improved decisionmaking**	Savings associated with improved decisionmaking because of improved data quality and access.
Evaluation	Analysis and reporting costs associated with determining whether the system is meeting user needs and organizational expectations.	**System security**	Savings associated with decreased risks to an organization's data (e.g., improved identification of sensitive/confidential data permits focused security efforts).

* In addition to readily measurable benefits, less quantifiable benefits, sometimes called "soft" or "intangible," occur as well. Examples include improved data use to keep more students in school, improved staff morale because employees trust the organization to maintain accurate human resources files, and more effective auditing procedures like error checking to confirm calculations. While it is difficult to place a dollar value on these "soft" benefits, they are very real and could be estimated and reasonably included in a cost-benefit analysis.

Net cost = Sum of benefit savings - sum of implementation costs

$$\text{ROI} = \frac{\text{Total cost savings - total tost of ownership}}{\text{Total cost of ownership}} \times 100$$

quality and use, and help users better understand the data they are analyzing. This can lead to savings associated with improved decisionmaking. Some financial implications of improved decisionmaking can be estimated as well. For example, a better command of information may improve purchasing choices; staffing decisions; and even academic preferences, such as curriculum selection, teaching assignments, and leadership decisions.

Exhibit 4.5 presents an example of several frequently recognized categories of costs and benefits (including cost avoidance) and the return on investment (ROI) that can accompany metadata solutions. Note that these categories are illustrative and may vary for individual organizations depending on a wide range of factors. These costs and benefits can be placed in a spreadsheet with columns to estimate dollar values.

Build-Versus-Buy Analysis

Deciding whether to build or buy a metadata system can be a challenge. Starting from scratch without being sure you have the human resources needed to handle the job can be overwhelming, but commercial products bring their own limitations. For example, most commercial packages are proprietary and cannot be modified without invalidating warranties and, in some cases, preventing upgrades from working properly. In addition, many commercial systems are designed with a very limited set of metadata that cannot meet even the most basic information needs common to education organizations. A "build or buy" choice has other ramifications as well, often including dictating whether metadata system architecture reflects centralized, federated, or distributed architecture (see below). Answers to the following questions can help planners decide whether to build or buy a metadata system.

Standard metadata components in many off-the-shelf products tend to be quite basic. The purchased systems generally must be customized if they are to meet the unique needs and circumstances of the organization and its existing data system; such customization may invalidate warranties and prevent upgrades from working properly.

- ✓ *Have other organizations with comparable needs and budgets found acceptable commercial solutions?* If so, perhaps those technology solutions might work for your organization as well. If not, an off-the-shelf product may not work for your organization either.
- ✓ *Do commercially available products meet all of your organization's needs or will they need to be modified?* If a product meets most, but not all, of your requirements, you may wish to determine whether it is possible to modify it, or reconsider the importance of any unmet needs. Adding to, or changing, a proprietary product's existing functionality is sometimes feasible; but modifications to improve processing speed or other aspects of performance, or to enable it to run on different platforms, are usually not. In addition to potentially invalidating warranties, customization of commercial products often causes the modified tools to become incompatible with future releases or updates from the developer. Before proceeding, confirm that support will still be provided even after you have modified the product.
- ✓ *Will commercially available products accommodate change over time?* Policies, business rules, and metadata characteristics are not constant. Priorities and procedures occasionally change, and a metadata system must be able to accommodate these changes when they occur.
- ✓ *Do you have access to staff or consulting resources with the necessary expertise to build your system?* If so, is the allocation of staff time or the cost to hire outside expertise within the range of the resources you have for the project? If you must hire outside expertise, have you determined how your staff would support a system they did not develop?

✓ *Do you have resources to support the system in an ongoing way?* Have you planned for ongoing costs such as new staff member training, system upgrades, and ongoing licensing? Whether building or buying, initial development costs, as substantial as they might appear, are not the only resources needed to maintain a system over time.

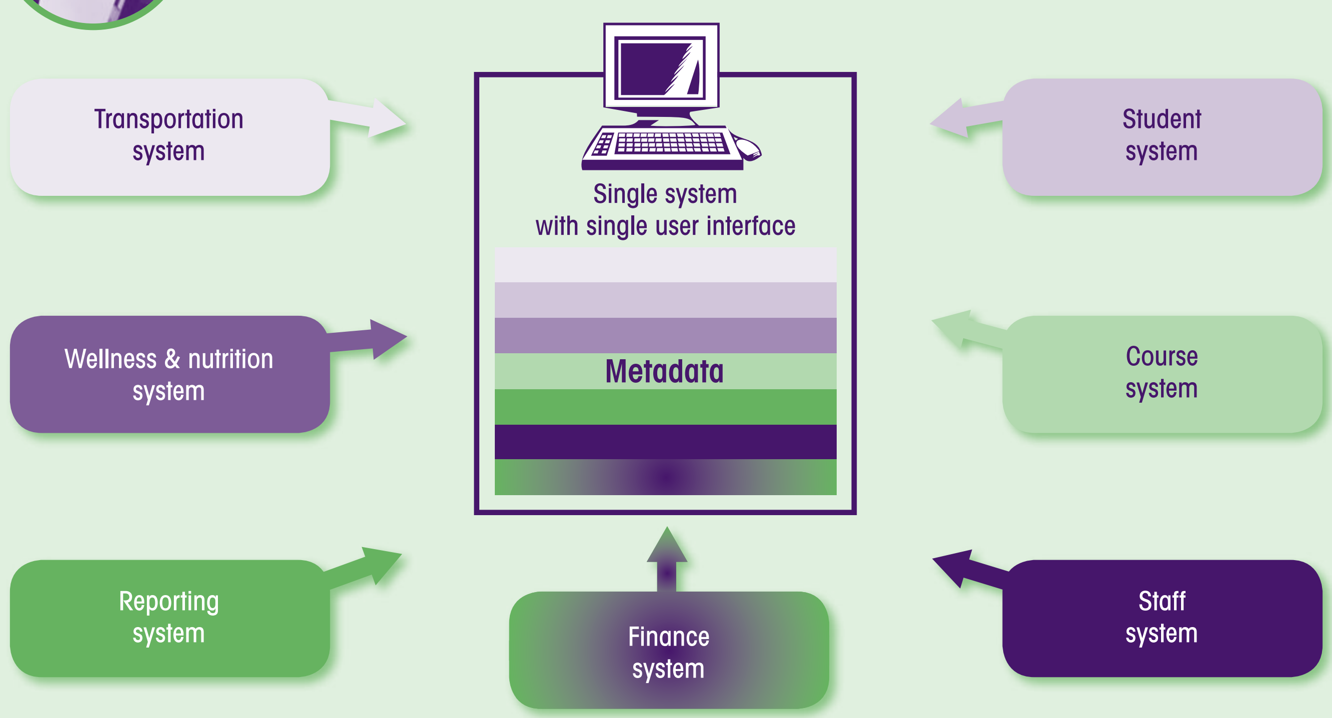
Metadata System Architecture

Metadata system architecture often is driven by the results of a build-versus-buy analysis that, in turn, depends on the organization's existing management, governance, and technology considerations. In the broadest sense, metadata system architecture can be divided into three main designs: centralized, federated, and distributed.

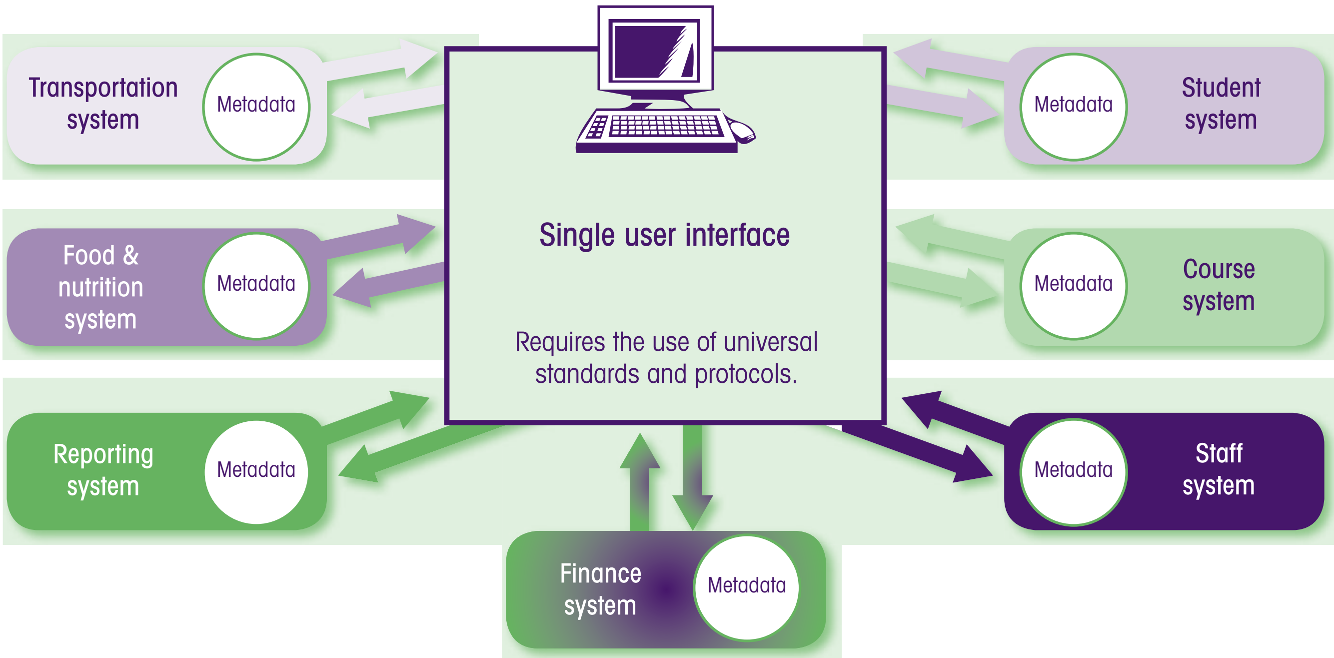
Centralized Architecture: As one might expect of a "centralized" system, all metadata exist in a single database that stores nothing but metadata (see exhibit 4.6). The greatest challenge to implementing a centralized architecture is finding a single model that meets the needs of all data systems and users. If a single metadata model has been designed for the entire organization, a centralized metadata system generally is fairly straightforward to implement. Centralized systems are governed, managed, and operated as a single entity. In other words, decisionmaking is also largely centralized, which helps ensure metadata are consistent across subsystems throughout the entire organization—for example, the definition and attributes of "class" would be the same in the finance system as in the student record system. Data stewards and data users generally access a centralized metadata system via a single interface, although the core interface may be modified to accommodate differences in access privileges or other user rights.

Exhibit 4.6

Centralized metadata system architecture



Federated metadata system architecture

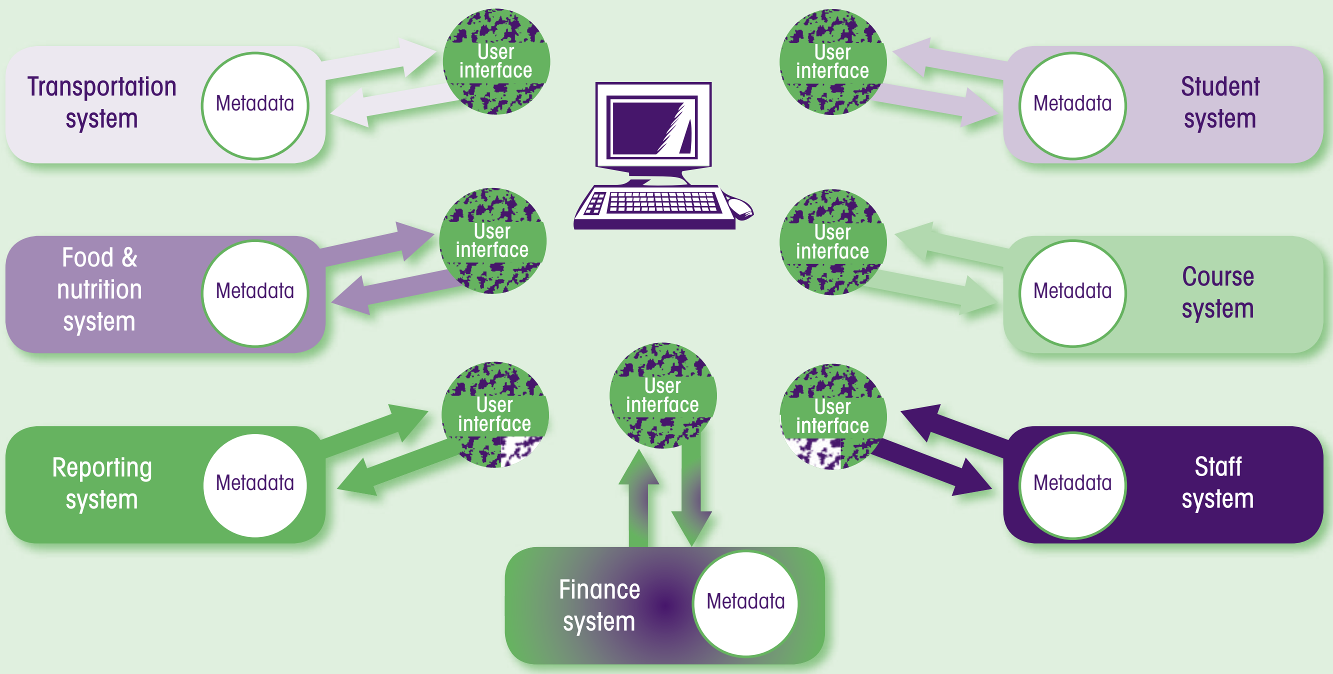


Federated Architecture: In federated designs, each stand-alone data system in the organization maintains its own metadata system within the constraints of a centralized technical framework and governance structure. This allows metadata to reflect the specific information needs of each independent data system while still ensuring communication capabilities with other independent systems. Users who access multiple data may do so through separate interfaces, and data stewards likely manage each system independently. However, metadata items that affect more than one system can be coordinated through automated translation and update processes, or by manual modification. Because of this, federated design requires central planning and rulemaking within a distributed architecture (see exhibit 4.7) and demands a fairly sophisticated technical infrastructure and strong system governance.

Distributed Architecture: In a distributed architecture, each stand-alone data system has a corresponding stand-alone metadata system. The major benefit of a distributed system is that metadata can be modified and updated without the need to coordinate with other systems (see exhibit 4.8). While there are other benefits to distributed architecture (for example, metadata directly reflect related operational data), cohesiveness and integration are generally lacking and stand-alone components tend to evolve without adherence to universal rules and conventions that permit synchronization with the rest of the system. Moreover, vocabularies and definitions often "drift," or start to deviate from those in other systems, usually leading to multiple terms for one item and, conversely, multiple items referenced by the same term. In either case, duplication arises and data quality suffers. Given such drift, these stand-alone components, sometimes called "silos," can become autonomous and independent over time, and eventually unable to exchange data or otherwise work with the rest of the system.

Exhibit 4.8

Distibuted metadata system architecture



Establishing a Project Implementation Plan

A thorough and realistic project plan is critical to an efficient and effective metadata system implementation effort. Planners should recognize the iterative nature of developing and implementing a technology initiative as complex as a metadata system, and budget the time needed to plan, implement, test, and refine the system repeatedly until user requirements are met. The implementation plan and schedule should address all aspects of the project, from planning through post-implementation training. Good plans often

- start with a basic and understandable feature that stakeholders are likely to care about, rather than a component that may be important, but does not speak clearly to user needs or experiences;
- build in evaluation time for a "feedback loop" that supports the iterative nature of developing and implementing a technology initiative; and
- stress extensibility, which allows modules to be expanded or customized after initial implementation has been successful—in other words, once stakeholders have mastered the basics, functionality can be extended to include more specialized capabilities.

A development schedule is only effective if its goals and deadlines are realistic. If they are unattainable and targets are missed, subsequent deadlines lose their credibility.

A project implementation plan should present work in discrete, manageable tasks. For example, mapping a metadata item inventory to all active data elements in a large education data system is a very big job—potentially too big to be accomplished in a single step. Instead, more manageable tasks might be identified and prioritized, such as mapping a smaller set of core metadata items. Another approach might be to divide a large, systemwide mapping job into subtasks based on data categories: student personal information, student enrollment, student assessment, staff personal information, staff assignments, etc. The activities in the implementation plan are then assigned, carried out, monitored, and completed in discrete units that can be understood and undertaken by members of the implementation team.

One special consideration when implementing metadata systems is the coordination required between metadata and existing (or envisioned) data systems. If an organization does not have a clear understanding of its current data–what exists, what format data are in, where they are located, and what quality they are–a metadata system that is dependent on those data is unlikely to provide useful information.

Tried and true tips for developing an implementation schedule

- ✓ Reduce large tasks to more manageable subtasks to keep jobs achievable.
- ✓ View the first attempt at a task that must later be repeated as a pilot effort. Learn from the experience and modify subsequent efforts (and timelines) to reflect lessons learned.
- ✓ Phase in functionality rather than triggering an all-or-nothing event. This may require more time initially, but will reduce wasted effort in the long run when lessons learned in early phases improve subsequent decisionmaking.

Training Users to Maximize System Utility

Metadata is not an inherently well-understood topic, and many stakeholders may not yet likely to be familiar with the term. Thus, professional development must be provided to system users. In many environments, including education, readily available data tools are not used to their full potential because ineffective or insufficient training makes using the system more of a challenge than a benefit. Metadata system training requires commitment from the organization to identify or develop skilled trainers, customize training curricula to reflect specific user needs, and allocate professional development time to the full range of stakeholders upon initial system implementation and ongoing use. In addition, metadata training efforts can be challenging because, unlike other technology initiatives, in most cases the organization's stakeholders have not asked for the system because they do not yet understand their need for this powerful information management tool. Without comprehensive training, it is extremely unlikely that stakeholders will appreciate the power and benefits of a metadata system.

Even the best designed metadata system will not work well if the people expected to use it don't understand its purpose or how to operate the system effectively.

The primary purpose of stakeholder training is to teach users to: (1) understand the purpose of metadata; (2) operate a metadata system effectively and efficiently; and (3) use metadata to inform their data use. Unless these major objectives are accomplished, only technical staff may have the confidence to use the metadata system and its potential value will never be realized.

Important considerations when planning a training program include

- *Introducing the concept of metadata.* Different stakeholders will have a widely varying understanding of metadata. Therefore, training programs should be designed so that those unfamiliar with the concept will not be overwhelmed with technical details, while anyone with some familiarity will not be bored. One strategy for providing this type of customized training is to adopt a modular approach, with each module building on content from the previous session. Stakeholders can begin their training at the level most appropriate for their knowledge and experience. The initial training module might, for example, introduce the concept of metadata without delving too deeply into technical details and terminology. A subsequent module could begin to address more formal terminology and model relationships between metadata, data, and information needs. A third module might then describe the organization's preferred practices for entering, managing, and using metadata.

- *Including meaningful, "real" examples to illustrate training points.* Participants in training activities generally appreciate lessons that can be readily applied to their everyday responsibilities. Training is most meaningful when it is clearly applicable to the participants and their jobs. Good trainers often illustrate points with "real life" examples that are directly related to the duties of the participants. In addition to explaining concepts in understandable terms, examples demonstrate how to use metadata "on the job," and they effectively illustrate metadata's power to improve data use.
- *Customizing training to match audience needs.* Not all stakeholders will use metadata the same way. For example, data stewards generally will be responsible for entering and updating most nontechnical metadata, whereas database administrators often are in charge of technical metadata. Program staff and other data users, on the other hand, need to focus on accessing metadata to improve their analysis and use of program data. Because each stakeholder group may use a metadata system in a slightly or substantially different manner, it often makes sense to develop separate training modules that can be combined as appropriate to meet the needs of each major user group. Customizing content to meet functional needs and minimize less relevant information generally makes training efforts more efficient and effective.

Teaching metadata in a training program

Effective training sessions often begin with ideas that stakeholders understand and then proceed to more advanced topics.

- ✓ What are metadata?
- ✓ How do metadata affect you and your data use?
- ✓ Why does the organization need metadata?
- ✓ Metadata system overview
  - Access rights and tools
  - Governance
  - Policies and procedures
- ✓ What are the basic (or advanced) system components and how are they accessed?
- ✓ How will metadata affect a user's understanding of data?
  - Data element definitions
  - Permitted values
  - Usage guidance
  - Restrictions
- ✓ Usage examples (related to audience)
- ✓ How are users expected to maintain system security?
- ✓ How do stakeholders learn more about the metadata system?

See appendix D for a description of a metadata training program.

Metadata will be a new concept to many participants. Training stakeholders to use a metadata system does not necessarily ensure they understand when or why to use it. In addition to describing the concept of metadata, trainers need to explain why metadata are relevant to each stakeholder group's roles and responsibilities.

- ✓ Policymaking staff might be shown how metadata can provide access to data usage instructions, term definitions, and interpretation guidance that will ensure their policy decisions are based on an accurate understanding of the data. Or, they might learn how the data are commonly used, and the implications of mistakes in collection and processing.
- ✓ Technology staff might be taught that metadata will provide a clear list of technical attributes (e.g., data element type and field length) that do not need to be reconsidered each time an item is collected. They might also need to learn that metadata can identify sensitive/confidential data and improve system security, or that metadata will simplify the exchange of data between systems, both within and outside the organization.
- ✓ Program staff might learn how metadata can help identify redundant data elements and collections, potentially reducing collection demands and improving data comparability and continuity over time. They might also learn that metadata can improve data checking and auditing to increase the overall quality of the data.

Regardless of the examples used, stakeholders should leave a training session with a clear sense of what metadata are and why using them is worth their time and effort. See appendix D for a detailed outline of a metadata staff training program.

Don't assume that stakeholders understand the power and possibilities of metadata—teaching them how and, sometimes more importantly, why to use a metadata system are critical aspects of any implementation effort.

The Key Points of This Chapter…

- Metadata systems are built around existing data systems and, ideally, the organization's vision for future data use and management. Introducing a metadata system is a complex project that requires planning efforts comparable to any other major initiative. Unlike other technology initiatives, however, an organization's stakeholders in most cases will not have asked for a metadata system because they may not yet appreciate its potential for helping them better understand and use data.
- Issues specific to implementing a metadata system include
  - building capacity given that most organizations do not have sufficient metadata expertise within staff ranks;
  - identifying metadata needs even if stakeholders are not familiar with the concept and potential of a metadata system;
  - justifying metadata through strict cost-benefit analysis given that many of the benefits of a metadata system cannot be quantified by traditional market pricing;
  - determining whether to choose an off-the-shelf product given that a metadata system must be customized to meet each organization's unique data needs and user requirements;
  - determining whether to pursue a centralized, federated, or distributed system architecture;
  - establishing a clear understanding of the organization's existing data–what data exist, their format, their location, and their quality–before expecting a metadata system (which is dependent on those data) to provide useful information; and
  - training stakeholders to understand metadata as a concept before they use a metadata system as a tool for improving the management and interpretation of the organization's data.

Chapter 5

Conclusion

This chapter summarizes why it is imperative for education organizations to develop and implement a robust metadata system.

The comforting adage, "the data speak for themselves," is often untrue in real life (see exhibit 5.1). In the complicated world of education data, answers to even apparently straightforward questions often depend on complex data. The "simple" question in chapter 1 illustrates this point: *How many eighth grade English teachers are in your schools?* This type of inquiry prompts a prudent school leader to ask other, even tougher, questions, such as:

- ✓ Are data queries answered correctly and consistently in my organization?
- ✓ Would different staff members give the same (or a different) answer to the same question?

Some organizations rely on the experience of their data steward(s) as the primary source of information for understanding and interpreting their data. In these

Exhibit 5.1

Instantiation: One reason your organization needs metadata

The connections between data and metadata can be quite complex. One fairly common example of the complicated nature of data systems—and why metadata are so important—is the concept of "instantiation," which refers to how a piece of data may change in presentation or format without changing in meaning at various "instants" during its life cycle within a data system. For example, note the flow of gender data that might occur in a typical education organization's data system:

Collection format (on the form):	Sex = Boy or Girl
System entry interface (how the data are entered):	Sex = M or F
Data warehouse format (how the data are stored):	Sex = 1 or 2
Data report (how users see the data):	Sex = Male or Female

Such change in nomenclature for the same information throughout a single system requires a mechanism for data users and managers to identify element code lists at each instance, regardless of the changing source and target. A robust metadata system is the most systematic, reliable, and accurate way of accomplishing this vital data management task.

organizations, that staff member's mind is the metadata system—the resource that describes, explains, locates, or otherwise makes it possible to retrieve, use, or manage data. But in this era of data-driven decisionmaking, the sheer volume of data collected for administrative, instructional, and management purposes complicates data systems beyond the management capacity of even the most experienced professional. There are simply more data to organize, access, and understand than ever before; and no data steward's mind, however powerful, is up to the task of managing all that data about data. A metadata system is not only a better and more reliable alternative—it is the only realistic way to effectively accomplish this vital information management task.

Metadata systems are critical components of an effective information management system. The benefits of properly implementing a robust metadata system include

- ✓ improving the likelihood of meeting users' information needs;
- ✓ improving the efficiency of data access and integration;
- ✓ improving the probability of correct data interpretation and use;
- ✓ identifying what data exist (and their location) throughout an organization;
- ✓ identifying redundancy and disparity in data sets;
- ✓ increasing the efficiency of data storage and maintenance;
- ✓ improving the accuracy of data transfer across systems;
- ✓ improving the application of business rules and edit checks;
- ✓ reducing user expertise required to conduct effective queries;
- ✓ advancing data quality;
- ✓ ensuring the proper maintenance of information over time; and
- ✓ improving the quality of data-driven decisionmaking.

While metadata cannot eliminate every opportunity for improperly collecting, using, or reporting data, a sound metadata system provides a framework for better understanding data and, therefore, minimizes the likelihood of data misuse. Cost–benefit and return-on-investment analyses ensure that both positive and negative implications have been considered prior to making a significant investment in time and resources to introduce a metadata system. In fact, many organizations find that the potential improvements to data quality and use are well worth the costs of developing, implementing, and supporting a robust metadata system.

Nevertheless, despite the potential value of metadata systems, many organizations have yet to develop and implement robust metadata systems. Leaders in these organizations may make this decision passively (they do not think about it) or actively (they reject the notion as too costly). When education leaders make an intentional decision not to implement a metadata system, they often do so because they believe that developing metadata items and systems:

- ✓ demands expertise that their staff may not possess;
- ✓ involves a great deal of work;
- ✓ takes a lot of time;
- ✓ costs a fair amount of money;
- ✓ requires a thorough understanding of current data resources;
- ✓ may require correcting existing deficiencies in data quality; and
- ✓ involves a long-term investment that does not match short-term goals.

All of these reasons for not developing metadata systems are valid—up to a point. Developing a metadata system is a substantial undertaking that requires significant time, expertise, commitment, and money. But like other time-, staff-, and resource-intensive initiatives (e.g., installing new networking systems, building new facilities, and introducing new professional development programs), metadata systems should yield benefits that far outweigh their costs.

No matter what the anticipated benefits are, a decision on whether to proceed with a metadata system will eventually be based on the relative costs and benefits of the proposed system. Planners engage in cost–benefit analysis to ensure that both the positive and negative implications of a metadata system have been considered. In addition to many readily measurable benefits, less quantifiable benefits exist as well. These are sometimes called "soft" or "intangible" benefits, and include improved data use when, for example, teachers can identify potential dropouts; improved staff morale, such as employees trusting that the organization maintains accurate human resources files; and more effective auditing procedures, such as error checking to confirm calculations. While placing a dollar value on these "soft" benefits is difficult, they are nevertheless real and could be estimated for the purposes of cost–benefit analysis. Many organizations have found that the potential improvements to data quality and use are well worth the costs of developing, implementing, and supporting a robust metadata system (see exhibit 5.2).

Given the complexities involved in designing and developing a metadata system, an education organization cannot decide it needs metadata, develop and institute a system, and expect it to be operational that same day. A thorough and realistic project plan is critical to implementing the effort effectively and getting the job done efficiently. A robust metadata system will provide context for a single data item, serve as the backbone for efficient data management, and improve the use, analysis, and management of any body of data. The result will be improved accuracy, utility, and comparability of elementary and secondary education data at local, state, and national levels.

Exhibit 5 2

Metadata: Solving the Case of the Inaccurate Dropout Count, Chapter 5

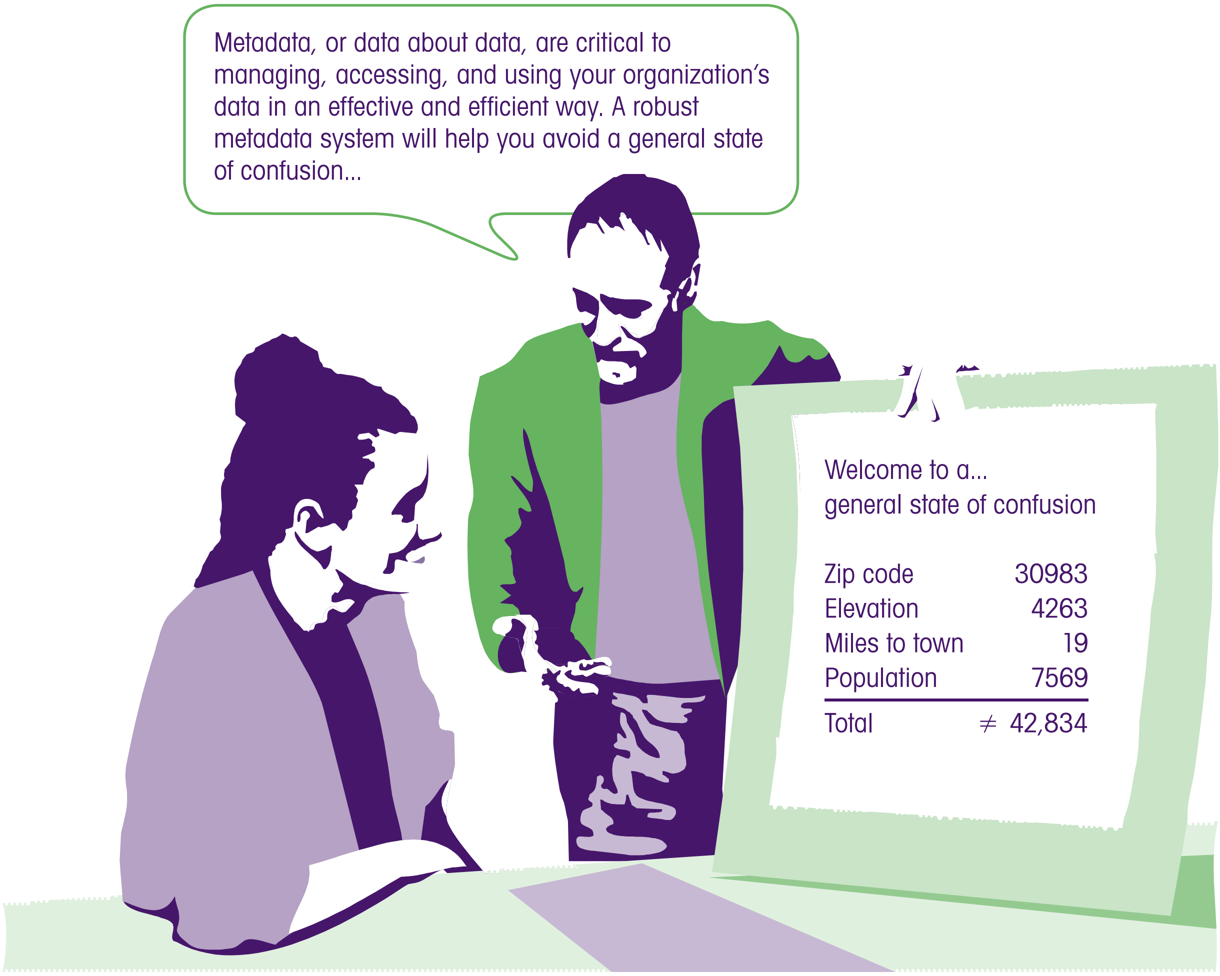
Continued from page 31.

Principal Howell and Mr. Olsen met Superintendent Sanders in her office the next morning to explain the error and submit revised dropout data for Lincoln High School. The superintendent reviewed the report and applauded the decrease in dropout rates, which reflected what Mr. Howell had assured her was occurring at the school. Despite her satisfaction, she became quite serious as she turned back to them. "So how is it that this critical information about the change in enrollment codes wasn't communicated effectively?"

Mr. Howell stammered for a moment, but Mr. Olsen was ready. "We need 'metadata,' " he replied in a measured tone. "A metadata system would, among other things, include definitions, code lists, and any changes in data items from year to year on the web form the schools use to report their data. Our data providers and users would have one-click access to the up-to-date guidance they need to understand what is being requested, and in what format. I know it sounds like something magical, but it's not. It is a standard practice for data collection, management, and use that is designed specifically to help avoid these types of data quality problems."

Dr. Sanders nodded. "I remember reading about metadata, but thinking that it seemed like an investment in the theoretical. Clearly this episode illustrates that there are practical implications of understanding our data better." Mr. Olsen smiled. "Don't get me wrong, there is an investment in instituting a robust metadata system," he offered cautiously. "Oh, I am sure," the superintendent agreed. "Starting with a rigorous training program! Evidently we haven't mastered the quality assurance procedures we already have in place. But such an investment would be a small price to pay if it means we can feel confident about using data to make sound management, instructional, and planning decisions."

The end


Metadata, or data about data, are critical to managing, accessing, and using your organization's data in an effective and efficient way. A robust metadata system will help you avoid a general state of confusion...
Welcome to a...
general state of confusion
Zip code 30983
Elevation 4263
Miles to town 19
Population 7569
Total ≠ 42,834

Appendix A. Common Word Lists for Naming Metadata Elements

This appendix provides an example of standard definitions for common words that can be adapted to meet the needs of state and local education agencies around the nation. This type of tool helps to standardize data element names throughout an organization.

An important initial step to consider when developing a data element is to adopt a set of common words (or lexicon) as naming conventions. This will ensure that data names are consistent and distinguishable throughout the organization. This appendix presents an excerpt of common words used by the California Department of Education to name and define commonly used terminology related to data.

Common word list for naming data elements: Data attributes

(Source: California Department of Education)

This table offers examples of commonly used terms and is not exhaustive.

Data attributes-Examples of common words	Common word definition
Amount	A monetary quantity in some type of currency, such as dollars. Not a count of capacity.
Code	Alphabetic or numeric coded data values that indicate the existence of a code table data entity. Should not be used if data values are not coded.
Comment	A textual comment about an object. Similar to Explanation.
Constant	A numerical value that does not change over time or circumstances and is used in one or more calculations.
Count	The number of objects that exist or have occurred. Not the same as Amount or Number.
Duration	A time interval or duration representing an elapsed time or the length of time an event lasted.
Explanation	A textual explanation about an object. Similar to Comment.
Flag	An indication that some event has happened or should happen. A binary situation, such as, on/off, true/false, 0/1, or yes/no. Similar to Indicator.
Identifier	A unique alphabetic identification of an object. Generally not just a number.
Indicator	An indicator of two possible conditions. A binary situation, such as, on/off, true/false, 0/1, or yes/no. Similar to Flag.
Value	A monetary value that indicates the worth of something in some type of currency. Not the same as Amount.

Common word list for naming data elements: Data entities

(Source: California Department of Education)

This table offers examples of commonly used terms and is not exhaustive.

Data entities-Examples of common words	Common word definition
Account	A data entity for tracking and managing monies, such as customer account.
Authority	A data entity for approval or delegation, such as expenditure authority.
Authorization	A data entity showing approval or delegation, such as employee account authorization.
Budget	A data entity for planned monetary or effort expenditure, such as yearly budget.
Category	A data entity representing a classification or grouping of objects, usually higher than class or group, such as equipment category. The hierarchy should be common for the organization, such as category, group, or class.
Class	A data entity representing a class of objects, such as facility class. The hierarchy should be common for the organization, such as category, group, or class.
Detail	A data entity containing details about a parent data entity, such as customer account activity detail.
Exemption	A data entity containing an exclusion or different situation, such as tuition exemption.
Expense	A data entity for expenses, such as goods or services, or for monies expended, such as equipment expense.
Group	A data entity representing a specific grouping of objects, such as building group. The hierarchy should be common for the organization, such as category, group, or class.
Summary	A data entity of "summary of data" above a detailed operational level, such as employee leave summary. Usually used in an executive information system (EIS).
Type	A data entity representing types of objects, such as equipment.
Validation	A data entity for verification or proof of a combination of objects, such as equipment part validation.

Appendix B. The Tale of the "No Pass-No Play" Business Rule

This appendix provides a real-world example of how a business rule gets developed.

Once upon a time, a school principal grew tired of his students' lackluster academic performance. Following staff complaints that students were giving less attention to their academics than their extracurricular activities, the principal decided to implement a "No Pass–No Play" policy. The next morning, he sat down and perfunctorily typed a brief memo, which was then distributed to every school teacher, coach, music instructor, and other extracurricular activity leader. It read simply:

"No Pass–No Play"

Thinking his work was done, the principal went on with his regular activities and awaited great improvements in the students' performance. But, the school staff were confused. What did the students need to pass, exactly? Homework assignments or test scores? For the week, month, or semester? Was this rule just for sports, or for band and chess club as well? The staff decided to ask for clarification. Hearing their concerns, the principal realized how ambiguous the memo had been and assured them that he would resolve the issue.

To avoid needless denial of play time to students exhibiting acceptable academic performance, yet prohibit extracurricular participation to those with subpar grades, the principal converted the imprecise rule into a more sophisticated form: a set of business rules. After some reflection and hard work, he sat down and wrote a new memo. It read as follows:

- ✓ No student may play unless he/she passes.
- ✓ A "student" is any child enrolled in, or provided education services by, a school.
- ✓ "Play" refers to engagement in any school-administered athletic, academic, musical, artistic, or other type of after-school activity.
- ✓ "Pass" means to achieve a grade of 70 or higher in every class in each school marking period. If a failing grade is awarded to the student for at least one class in a marking period, he or she is prohibited from "play" for the entirety of the following marking period. This includes the summer break.
- ✓ No exceptions.

This more detailed presentation of the rule clarified both the intent and the administration of the No Pass–No Play rule.

Pleased with the new set of business rules, the principal decided it would be a good idea to start collecting data about the school's No Pass–No Play activities. In addition to the student data the school already collected, information about each student's No Pass–No Play status would also be gathered. Teachers and leaders of extracurricular activities would be responsible for collecting and reporting the new data, and administrative support staff would enter them into the school's record system. Two new data elements were to be included:

1) "All Grades Passed?" (Did the student pass all of his/her classes?)
2) "Extracurricular Participation?" (Did the student participate in any after-school extracurricular activities?)

However, to facilitate proper and consistent coding into the data system, additional rules were needed:

- ✓ Data about students' academic performance must only be reported by their teachers.
- ✓ Data about students' extracurricular participation must only be reported by leaders of those activities.
- ✓ The value for the new data element "All Grades Passed" must be either "Yes" or "No," without exception, and must not be "Null" or left blank.
- ✓ The value for the new data element "Extracurricular Participation" must be either "Yes" or "No," without exception, and must not be "Null" or left blank.
- ✓ The value for "Extracurricular Participation" may be "Yes" only if the value for "All Grades Passed" is "Yes."

These business rules reflected both an operational decision (if you don't pass, you can't play) and the guidance needed to accurately collect data related to that policy choice. And because these new business rules were clear and unambiguous, the school achieved better academic performance and the principal, staff, and students all benefited.

In this example, a basic policy was converted into a business statement, and then into a set of actionable business rules useful in operations and the creation and maintenance of a data system. The policy, formulated in the principal's mind, had a general goal of improving academic performance by allowing only those students with passing grades to participate in extracurricular activities. This policy was then converted into a more specific, yet still overly general, business-rule statement expressed in the original memo as, "No Pass–No Play." Subsequently, that statement was translated into two sets of atomic, executable business rules intended to realize the policy's goal. These business rules were then used to improve school operations and student performance, as well as direct the creation of new data elements and their maintenance in the school's data system.

Appendix C. Metadata Standards and Registries

This appendix defines the concept of a metadata registry and includes descriptions of several well-known metadata registries and standards available online.

Metadata registries are centralized locations where metadata structures and definitions are stored and maintained, often online. The following is a brief description of several well-known metadata registries.

NCES Handbooks Online

One widely recognized data standard in the field of elementary and secondary education is the *NCES Handbooks Online*, which defines standard education terms for students, staff, schools, local education agencies, intermediate education agencies, and state education agencies. It is intended as a reference document for public and private education institutions and early childhood centers, as well as education researchers and other users of education data. This web-based tool allows users to view and download information about data elements via an electronic table of contents, a drill-down finder, element-name and first-letter searches, and advanced query options. Visit *http://nces.ed.gov/programs/handbook* for more information.

The SIF Implementation Specification

Another widely respected standard for education data is the SIF Implementation Specification. Developed by the Schools Interoperability Framework Association (SIFA), this is a set of rules and definitions that enable software programs from different vendors to share data without additional programming by local schools, districts, or states. SIFA's goal is to allow school administrators, teachers, and other school staff to have secure access to the most current and accurate education data available. Visit *http://www.sifinfo.org* for more information.

Additional Registries

Other sets of standards more specific to metadata systems include metadata registries. These are centralized locations where metadata structures and definitions are stored and maintained. Numerous nationally and internationally recognized metadata registries are available online, including

- ✓ The Dublin Core Metadata Initiative is developing interoperable online metadata standards that support a broad range of purposes and business models. Visit *http://dublincore.org/dcregistry* for more information.
- ✓ The National Information Exchange Model (NIEM) is designed to develop, disseminate, and support enterprise-wide information exchange standards and processes that enable jurisdictions to effectively share critical information, and

support the day-to-day operations of government agencies throughout the nation. Visit *http://www.niem.gov* for more information.

✓ The International Organization for Standardization (ISO) metadata registry consists of a hierarchy of data elements, value domains, data element concepts, conceptual domains, and classification schemes that address the semantics of data, the representation of data, and the registration of the descriptions of that data. Visit *http://www.iso.org* for more information. (Note: Because "International Organization for Standardization" would have a different acronym in different languages, the organization decided that whatever the country or language, the short form of its name would always be ISO.)

Appendix D. Example of a Metadata Training Program

This appendix provides a description of a metadata/business intelligence staff training program that can be adapted to meet the needs of state and local education agencies around the nation.

Sample Metadata Training Program

(Adapted from the *Kansas Department of Education Data Project Training Strategy: Metadata and Business Intelligence Tools, Version 1.0.*)

1 Introduction

1.1 Scope of Document

This document outlines the training strategies and training timelines for an education agency's metadata and business intelligence (BI) system. This training consists of both metadata and BI content.

1.2 Background

Metadata, or "data about data," are crucial to an education data system because they drive the BI system applications, data marts, and interfaces that internal and external data consumers use for reporting, research, analysis, and decisionmaking purposes. The metadata provide data consumers with answers to the following questions:

- ✓ What data do I have?
- ✓ What do they mean?
- ✓ Where are they?
- ✓ How did they get there?
- ✓ How do I get them?
- ✓ How do I use them?

All users who access data will need instructions on what metadata are and how they can be used to produce meaningful data sets. Consequently, it will be necessary to train stakeholders on both: 1) metadata items so that they understand what data they have and what they mean; and 2) the metadata system so that they can access metadata to answer their business and research questions. The depth, content, and delivery media for this training will be dictated by the needs of the different user groups.

1.3 Purpose of Training

The purpose of this training is to familiarize end users with the concept of metadata, to demonstrate the relationship between metadata and the organization's data, and to teach end users to extract, format, and analyze data through the use of the metadata system applications, interfaces, and ad-hoc queries.

1.4 Scope of Training

Training content will initially be directed at internal stakeholders and will focus on the data warehouse and its associated metadata. The preliminary list of metadata elements that will be stored in the data warehouse is as follows:

- ✓ data element name;
- ✓ definition;
- ✓ business rule;
- ✓ permitted values;
- ✓ NCES definition;
- ✓ DW data type (string/number, length);
- ✓ transformation (crosswalk);
- ✓ previous source (server, db, table, column);
- ✓ data owner/steward;
- ✓ subject area/keyword;
- ✓ policies;
- ✓ data quality metrics;
- ✓ audit trail; and
- ✓ where used (e.g., Education Data Exchange Network).

2 Training Strategies

2.1 Course Modules

All training courses and materials will be pilot-tested internally prior to general release or use. The initial metadata and business intelligence training courses will include

- ✓ **Metadata 101** – an introduction to the concept of metadata and how it is used in daily life;
- ✓ **Using the Organization's Metadata System to Access Metadata** – a hands-on overview dedicated to understanding the contents of the metadata system;
- ✓ **Using the Organization's Metadata System for Data Stewards** – a variation of the previous course focused on entering, maintaining, and using the metadata and business intelligence system;
- ✓ **Business Intelligence System Training** – an introduction to the organization's BI system applications; and
- ✓ **Introduction to Metadata and the BI System Applications** – a hands-on course that combines the Metadata 101 course content with practice using the BI system applications.

2.2 Metadata and BI system Core Team

Most members of the metadata and BI system implementation team will already be familiar with the creation and use of metadata. As a result, this group will be the primary pilot test audience for all metadata and BI system training courses and materials.

2 2 1 Suggested courses

- ✓ Metadata 101
- ✓ Using the Education Agency's Metadata System
- ✓ BI System Training
- ✓ Introduction to Metadata and the BI System Applications

2 2 2 Training content

Training content for core metadata and BI system team members will include

- ✓ a description of the metadata being stored in the metadata system;
- ✓ the relationship between the source systems and the data warehouse metadata;
- ✓ a demonstration of how to access the data warehouse metadata;
- ✓ a discussion of the components of metadata (use, lineage, history, location in source system, etc.);
- ✓ a discussion of how to use metadata to assist with research and reporting;
- ✓ a demonstration on using the tool for entering, editing, and maintaining metadata; and
- ✓ hands-on practice with the BI system applications and interfaces.

2 2 3 Training delivery platforms

Delivery platforms for providing metadata and BI system training to the core team include

- ✓ presentations and hands-on workshops during scheduled metadata and BI system project meetings; and
- ✓ summary documents and project status emails.

2 2 4 Ongoing support

The training needs of the metadata and BI system core group will be monitored by the Project Director and Project Coordinator. Ongoing training and support will be provided on an as-needed basis.

2.3 Internal IT/Software Development Staff

IT staff and programmers require an overview of the metadata and warehouse in order to ensure the integrity and consistency of their programming code and web applications.

2 3 1 Suggested courses

- ✓ Metadata 101
- ✓ Using the Education Agency's Metadata System to Access Metadata

2 3 2 Training content

The training for IT and software development staff will include

- ✓ a discussion of the role metadata play in the education agency's metadata and BI system;
- ✓ review of the metadata associated with the BI system;
- ✓ a demonstration on how to access metadata in the metadata system;
- ✓ a demonstration on how to identify data's characteristics, description, uses, lineage, and history;
- ✓ a discussion concerning why metadata matter and how they may affect work; and
- ✓ hands-on practice using the system to access metadata.

2 3 3 Training delivery platforms

Delivery platforms for providing metadata training to the IT/programming staff include

- ✓ face-to-face presentations (at programmers' meetings, ad-hoc meetings, etc.);
- ✓ hands-on training with the metadata system;

- ✓ printed manuals; and
- ✓ printed quick-reference documents.

2.3.4 Ongoing support

Face-to-face refresher seminars will be held periodically throughout the year as needed.

2.4 Education Agency Data Stewards

Education agency data stewards have already been trained to create and maintain metadata. However, as the metadata and BI system continue to develop, all data stewards will need to know what metadata are stored in the data warehouse, how to update and manage metadata, how metadata relate to the data and metadata in their source systems, and how metadata can be used for research and reporting.

2.4.1 Suggested courses

- ✓ Metadata 101
- ✓ Using the Education Agency's Metadata System for Data Stewards
- ✓ BI System training

2.4.2 Training content

Training content for data stewards will include

- ✓ a description of the metadata being stored in the metadata system;
- ✓ the relationship between the source systems and the data warehouse metadata;
- ✓ a demonstration of how to access the system metadata;
- ✓ a discussion of the components of metadata (use, lineage, history, location in source system, etc.);
- ✓ a discussion of how to use metadata to assist with research and reporting; and
- ✓ hands-on practice entering, editing, and maintaining metadata in the metadata system.

2.4.3 Training delivery platforms

Delivery platforms for providing metadata training to the data stewards include (but are not limited to)

- ✓ face-to-face presentations (some of which occur in scheduled data steward workgroup meetings);
- ✓ hands-on training;
- ✓ metadata and BI system page of the education agency portal;
- ✓ printed manuals;
- ✓ printed quick-reference materials; and
- ✓ workbooks (with activities).

2.4.4 Ongoing support

Ongoing support will be provided through the monthly data steward workgroup meetings, periodic refresher presentations, the metadata and BI system webpage, and updated printed materials. Train-the-trainer sessions may be initiated to encourage participants to train their staff on the BI system and metadata.

What are Metadata and why do they matter? An everyday example

An everyday example of metadata use can help to explain the concept to stakeholders:

A temperature measurement (e.g., 28) is a piece of data, but to correctly interpret it as information requires additional context, or "metadata," which are data about data.

- What is the scale? Degrees Celsius or degrees Fahrenheit?
- What is the location? At the airport, downtown, or outside your window?
- What is the altitude of the measurement? At ground level (where it gets hotter and colder) or from a weather balloon a thousand feet above you?
- Is it the high or low temperature for the day?
- Is it the current temperature, yesterday's temperature, or tomorrow's predicted temperature?

This context is critical to using the piece of data as information. In its absence, 28 could mean 28° Fahrenheit (below freezing) in the middle of a parking lot as the day's high temperature (brr, it is cold). Equally plausible, however, it could mean 28° Celsius (or 82° Fahrenheit) in the shade on a grassy field as the day's low temperature (wow, is it ever a scorcher). This additional information that provides the "context" for understanding what 28 means is an example of metadata, and illustrates why "data about data" are necessary to derive meaning.

2.5 Education Agency Program and Support Staff

Education agency program and support staff will require training on metadata and the BI system applications for research and reporting purposes.

2 5 1 Suggested courses

- ✓ Metadata 101
- ✓ Using the Metadata System to Access Metadata
- ✓ BI System Training

2 5 2 Training content

Training content for program and support staff will include

- ✓ introduction to the concept and uses of metadata;
- ✓ description of the system's metadata and warehouse;
- ✓ the relationship between the source systems and the data warehouse metadata;
- ✓ demonstration of how to access the system's metadata;
- ✓ discussion of the components of metadata (use, lineage, history, location in source system, etc.);
- ✓ discussion of how to use metadata to help with research and reporting; and
- ✓ demonstration of the use of the BI system applications.

2 5 3 Training delivery platforms

Delivery platforms for providing metadata training to program and support staff include

- ✓ face-to-face presentations;
- ✓ hands-on training;
- ✓ metadata and BI system page of the education agency portal;
- ✓ printed manuals; and
- ✓ printed quick-reference materials.

2.5.4 Ongoing support

Ongoing support will be provided through the metadata webpage and printed materials.

2.6 Education Agency Executives

Education agency executives will require summary information about metadata and the BI system applications and how they can be used. Intensive hands-on training is not planned for this user group.

2.6.1 Suggested courses

- ✓ Metadata 101
- ✓ BI System Training

2.6.2 Training content

Training content for executives will include

- ✓ definition of metadata;
- ✓ how metadata fit into the greater metadata and BI system project;
- ✓ how metadata can be used internally and externally;
- ✓ description of how metadata and the BI system applications relate to one another; and
- ✓ description of the BI system applications and examples of the types of business questions that can be answered using data marts and BI system applications.

2.6.3 Training delivery platforms

Delivery platforms for providing metadata and BI system training to education agency executives include

- ✓ demonstrations posted on the system webpage;
- ✓ printed Q&A reference guides; and
- ✓ face-to-face demonstrations

2.6.4 Ongoing support

Ongoing support for education agency executives and administrators will be supplied through printed or emailed status reports and updates. Face-to-face or hands-on training will be available as needed.

2.7 School and District Administrators

School administrators will use the BI system applications to review high-level, summary data (such as their AYP status, QPA status, and general school/district information). Training on the metadata that drive the BI system applications will be embedded in their BI system training.

2.7.1 Suggested course

- ✓ Introduction to Metadata and BI System Applications

2.7.2 Training content

Training content for school administrators will include

- ✓ introduction to metadata;
- ✓ introduction to the BI system;

- ✓ instruction on using the metadata system (including querying and report formatting and filtering); and
- ✓ instruction on using the BI system applications to answer specific business questions.

2.7.3 Training delivery platforms

Delivery platforms for metadata and BI system training to school and district administrators include

- ✓ regional face-to-face presentations;
- ✓ interactive television (ITV) presentations;
- ✓ regional hands-on training;
- ✓ printed manuals, FAQ sheets, and other support materials;
- ✓ conference calls (as needed); and
- ✓ presentations at relevant conferences and user group meetings.

2.7.4 Ongoing support

Ongoing support will be provided through periodic conference calls, Help Desk support and supplemental face-to-face (hands-on) training provided on an as-needed basis throughout the year. Users can also email questions about the BI system applications to a metadata and BI system support email account.

2.8 School Staff

School staff, including teachers, secretaries, and testing coordinators are expected to represent one of the largest populations of stakeholders. These users will rely on the BI system applications and data marts to obtain information about data standards, licensure status, assessment results, and related information. Training on the metadata that drive the BI system applications will be embedded in their BI system training.

2.8.1 Suggested course

- ✓ Introduction to Metadata and BI System Applications

2.8.2 Training content

Metadata and BI system training content for school staff members will include

- ✓ introduction to metadata;
- ✓ introduction to the BI system applications;
- ✓ instruction on locating and using the BI system applications (including querying and report formatting and filtering); and
- ✓ instruction and discussion on using the BI system applications to answer specific business questions.

2.8.3 Training delivery platforms

Delivery platforms for providing training to school staff members include

- ✓ regional face-to-face presentations;
- ✓ ITV presentations;
- ✓ regional hands-on training;
- ✓ printed manuals, FAQ pages, workbooks (with activities), and other support materials;
- ✓ conference calls (as needed);
- ✓ demonstrations posted on the metadata and BI system webpage; and
- ✓ presentations at relevant conferences/user group meetings.

2 8 4 Ongoing support

Ongoing support for school staff will be supplied through conference presentations, periodic conference calls, Help Desk support, the metadata and BI system webpage, messages and updates on the metadata and BI system listserv, and supplemental hands-on training provided on an as-needed basis throughout the year. Train-the-trainer sessions may be initiated with this group in an effort to provide guidance as schools train their own staff on the use of the BI system applications.

2.9 Public

Members of the general public will eventually be able to use the BI system applications.

2 9 1 Suggested courses

No formal courses

2 9 2 Training content

Training content for the public will include

- ✓ introduction to the BI system applications;
- ✓ instruction on locating and using the system (including querying, report filtering, and formatting); and
- ✓ instruction on using the BI system applications to answer business questions.

2 9 3 Training delivery platforms

Delivery platforms for providing metadata and BI system training to the public include

- ✓ demonstrations posted on the data portal of the education agency website;
- ✓ printed quick reference guides and FAQ sheets; and
- ✓ podcast recordings of BI demonstrations.

2 9 4 Ongoing support

Ongoing support for the public will be provided through news bulletins and updates posted on the metadata and BI system and education agency web pages.

3 Example Timeline

3.1 Critical Milestones

Implementation of metadata training is dependent upon several tasks, including:

- ✓ metadata system developed (insert target date);
- ✓ metadata system test plan completed (insert target date);
- ✓ metadata application in production (insert target date);
- ✓ metadata system initial data load (insert target date);
- ✓ BI Capability Analysis Document (insert target date);
- ✓ selection of BI system suite (insert target date);
- ✓ initial BI products delivered (insert target date); and
- ✓ data marts developed (insert target date).

3.2 Training Timeline

The expected timeline for the metadata and BI system training is as follows:

Approximate timing	Training/deployment activities
Week 1	Metadata system is developed
Week 3	Metadata system test plan is completed
Weeks 3-9	Development of Metadata 101 and Using the Education Agency Metadata System training materials
Week 5	Metadata application in production
Week 6	BI solutions and matrix documents completed
Week 6	Metadata system initial data load completed
Week 7	BI Capability Analysis Document completed
Week 8	BI System application(s) selected and ordered
Weeks 9-10	Pilot testing of Metadata 101 and Using the Education Agency Metadata System training
Weeks 10-12	Updating Metadata 101 and Using the Education Agency Metadata System training materials, based on feedback
Weeks 12-20	Implementation of Metadata 101 and Using the Education Agency Metadata System training (internal staff)
Weeks 12-20	Development of BI System Applications and Introduction to Metadata and BI System Applications training materials
Weeks 16-20	BI System Products delivered and installed
Weeks 16-20	Pilot test of BI System Applications and Introduction to Metadata and BI System Applications training
Weeks 20-28	Updating of BI System Applications and Introduction to Metadata and BI System Applications training materials
Weeks 24-28	Data marts developed
Weeks 24-28	Implementation of BI System Applications training (internal)
Weeks 29-33	Implementation of Introduction to Metadata and BI System Applications training (external)

4 Training Evaluation

In an effort to ensure that training needs are met and to identify areas for curriculum improvement, quantitative and qualitative training feedback will be elicited from internal and most external participants via paper and electronic evaluation forms. The results of these evaluations will be aggregated by stakeholder group and compiled into summary documents that will be shared with the metadata core team. Adjustments will be made as needed to the training content, delivery mechanisms, and support materials.

Appendix E. Additional Resources

This appendix lists other resources related to metadata and education data quality, including sources referenced in the document and materials available from the National Forum on Education Statistics (the Forum), the National Center for Education Statistics (NCES), and other organizations.

These materials may be useful to school, district, or state education agency staff developing a metadata system.

References from Text

Clyde, A. (2002) Metadata, *Teacher Librarian, 30(2),* 45-47.

El-Sherbini, M. and Klim, G. (2004) Metadata and Cataloging Practices, *The Electronic Library,* 22(3) 238-248.

Lee, H., Kim, T., and Kim, J. (2001) A Metadata Oriented Architecture for Building a Datawarehouse, *Journal of Database Management,* 12(4), 15-25.

Michener, W.K., et al. (1997) Nongeospatial Metadata for the Ecological Sciences, *Ecological Applications,* 7(1), 330-342.

Shankaranarayanan, G. and Even, A. (2006) The Metadata Enigma, *Communications of the ACM (Association for Computing Machinery, Inc.),* 49(2), 88-94.

Other Metadata-Related Resources from the National Forum on Education Statistics

The National Forum on Education Statistics has produced a wide range of publications related to data quality and data management. These resources are available at no cost at *http://nces.ed.gov/forum/publications.asp.*

NCES Nonfiscal Data Handbook for Early Childhood, Elementary, and Secondary Education (2007)

Handbooks Online *http://nces.ed.gov/forum/pub_2003419.asp*

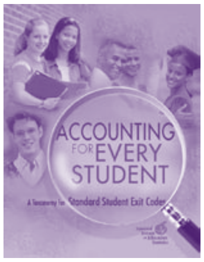
The *NCES Handbooks* are a valuable source of metadata for organizations and individuals interested in education data. These print and online resources define standard education terms for students, staff, schools, local education agencies, intermediate education agencies, and state education agencies. The handbooks are intended to serve 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 from the *Handbooks* via an electronic table of contents, a drill-down finder, element name and first letter searches, and advanced query options.

NCES Handbooks Online State Customization Tool (2007)

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

This is the content administrator site for *NCES Handbooks Online* (see above). The customization tool allows authorized state education agency staff to modify data elements, instances, and options for state use while still retaining the basic structure and organization of the *NCES Handbooks Online.* Access to this site requires a password.

Accounting for Every Student: A Taxonomy for Standard Student Exit Codes (NCES 2006–804)



http://nces.ed.gov/forum/pub_2006804.asp

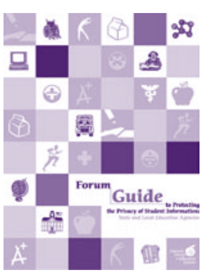
This "best practices" guide presents an exhaustive and mutually exclusive exit code taxonomy that accounts, at any single point in time, for all students enrolled (or previously enrolled) in a particular school or district. It is based on exit code systems in use in state education agencies across the nation and a thorough review of existing literature on the subject.

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 privacy interests of parents and students 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 confidentiality concerns related specifically to health-related information. Much of the guidance in this publication would be of interest to organizations generating business rules about privacy policies.

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

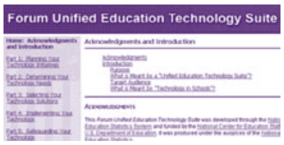


http://nces.ed.gov/forum/pub_2004330.asp

This 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. Much of the guidance in this publication would be of interest to organizations generating business rules about privacy policies.

Additional Resources

Forum Unified Education Technology Suite (2005)



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

This free online resource combines material from previously published NCES/Forum guides into one comprehensive document that will be updated periodically. The publication presents a practical, comprehensive, and proven approach to assessing, acquiring, instituting, managing, securing, and using technology in education settings.

Forum Guide to Building a Culture of Quality Data: A School and District Resource (NCES 2005–801)

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

Awareness has grown about the link between effective teaching, efficient schools, and quality data. The quality of information used to develop an instructional plan, run a school, plan a budget, or place a student in a class depends on the school data clerk, teacher, counselor, and/or school secretary who enter data into a computer. With that in mind, the focus of this report is on data entry—getting things right at the source.

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

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.

Links to Selected State Education Agency Data Dictionaries

Alaska: *http://www.eed.state.ak.us/data_mgmt/handbookuser/index.asp*

Colorado: *https://cdeapps.cde.state.co.us/DataDictionary*

Kansas: *http://kids.ksde.org under the Documents Tab*

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

Oregon: *http://www.ode.state.or.us/data/kids/datadictionary*

Utah's data warehouse documentation: *http://www.schools.utah.gov/warehouse*