Guide to the Education Data Model: Version 1 (PK-12)

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. 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.

January 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, PK-12 Data Model Task Force. (2009). *Guide to the Education Data Model: Version 1 (PK-12)* (NFES 2009-802). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

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 at 1-877-4ED-PUBS or order online at http://www.edpubs.org.

Technical Contact:
Kashka Kubzdela
(202) 502-7411
Kashka.Kubzdela@ed.gov

The Data Model Task Force

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

Task Force Members

Co-Chairs

E. Glenn McClain
Platte Valley School District, Weld Re-7 (CO)

Jeff Stowe
Arizona Department of Education

Members

Shawn Bay
eScholar

Pam Birkeland
Helena Public Schools (MT)

Bethann Canada
Virginia Department of Education

Les Daenzer
Lower Kuskokwim School District (AK)

Alicia Diaz
Southern Regional Education Board

Bertha Doar
Rockwood School District (MO)

Sonya Edwards
California Department of Education

Patricia Eiland
Alabama Department of Education

Larry Fruth
SIF Association

Ed Goddard
Clark County School District (NV)

Kathy Gosa
Kansas Department of Education

Stephen Hockett
US Office of Education Technology

Lee Hoffman
National Center for Education Statistics

Alex Jackl
ESP Solutions Group

Steve King,
ESP Solutions Group

Derrick Lindsay
Mississippi Department of Education

Blair Loudat
North Clackamas Schools (OR)

Stephen Metcalf
Montpelier School District (VT)

Lee Rabbitt
North Kingston School Department (RI)

Linda Rocks
Bossier Parish Schools (LA)

Robert Rodosky
Jefferson County Public Schools (KY)

Michael Strozeski
Richardson Independent School District (TX)

Charlene Swanson
New York State Education Department

Dave Uhlig
Charlottesville City Public Schools (VA)

David Weinberger
Yonkers Public Schools (NY)

Ray Wilson
Poway Unified School District (CA)

Connie J. Wise
Illinois State Department of Education

Raymond Yeagley
Northwest Evaluation Association

Principal Consulting Author

Vicente Paredes
Schools Interoperability Framework Association
Washington, DC

Project Officer

Kashka Kubzdela
National Center for Education Statistics
U.S. Department of Education
Washington, DC

Acknowledgements

The members of the Data Model Task Force of the National Forum on Education Statistics would like to thank all who contributed to the development of the design, content, tools, and documentation of the data model.

Much of the work of the Data Model Task Force was done in subgroups. The Forum would like to recognize the subgroup chairs listed below, who willingly accepted the burden. The Forum also thanks the members of the subgroups, many of whom were not members of the Data Model Task Force or the Forum.

Subgroup Members

School Formulation and Administration
Kathy Long (chair)
Steve King
Roger Young

Course of Study
David Weinberger (chair)
Patsy Eiland

Alternative/Supplemental Services and Instruction
Bethann Canada (chair)
Bettie Estes-Rickner
Dan Fandrey
Blair Loudat
Steve Metcalf
Ross Santy

Teaching and Learning
Raymond Yeagley (chair)
Victoria Bernhardt
James Boyd
Bertha Doar
Margaret Heritage
Alex Jackl
Gage Kingsbury
Derek Mitchell
Lee Rabbitt
Beth Young

Schools Improvement and Quality Management
Sonya Edwards (chair)
Nancy Burke
Les Daenzer
Alicia A. Diaz
Daniel Fandrey
Cathryn Huser
Steve King
Robert Rodosky
Charlene Swanson
Ray Wilson

Individual Student Tracking
Bertha Doar (chair)
Tos Crawford
Bruce Dacey
Patsy Eiland
Mary F. Gilmore-Dunn
Ed Goddard
Kathy Gosa
Sharon Hennessy
Linda Rocks
Kendra Shakir
Charlene Swanson
David Uhlig
Susan VanGorden
Connie J. Wise

The Task Force also acknowledges the support of the National Center for Education Statistics; Tim Magner and Hugh Walkup (Office of Education Technology) and Deborah Newby (Office of Planning, Evaluation, and Policy Development), U.S. Department of Education; Beth Young and Anthony Garofano of Quality Information Partners; Jill Abbott of the SIF Association; and, Nzinga Damali-Cathie and Amanda Miller of the Council of Chief State School Officers (CCSSO).

Additional thanks are due to the contributors to the Education Data Model, who participated in a special meeting on teaching and learning: Victoria Bernhardt (California State University), Kathryn Parker Boudett (Harvard Graduate School of Education), Margaret Heritage (University of California at Los Angeles, CRESST), Sharnell Jackson (Chicago Public Schools), Matt Howard (eTech Ohio), Jon Fullerton (Harvard Graduate School of Education), Gage Kingsbury (NWEA), and Erin McGoldrick (DC Public Schools).

Foreword

The National Forum on Education Statistics (Forum) is pleased to present this *Guide to the Education Data Model: Version 1 (PK-12)*. One of the main goals of the Forum is to facilitate the sharing of knowledge among experts from local education agencies (LEAs) and state education agencies (SEAs) nationwide in order to improve the quality of education data gathered for use by those making policy or program decisions about education. An approach to furthering this goal has been to pool the collective experiences of Forum members to produce "best practices" guides in areas of high interest to those who collect, maintain, and use data about elementary and secondary education. One area of high interest has been to develop a model of education data that could be used to improve interoperability and quality of, and access to data, as well as provide a standard for the meaning and use of data in education software systems.

In This Guide

- Chapter 1 discusses the objectives of this publication, explains how the Data Model was developed, and introduces the Data Model and additional supporting products.
- Chapter 2 provides an overview of and rationale for the Data Model, along with a discussion of the benefits and risks associated with the conceptual approach to data modeling used in this project.
- Chapter 3 discusses how to navigate the Data Model.
- Chapter 4 discusses the core concepts and content structure of the Data Model. It also introduces two open source tools for manipulating the Data Model.
- Chapter 5 provides examples for how to use the Data Model and its accompanying tools.
- Appendix A presents questions and answers about the Data Model.
- Appendix B is a glossary that defines key terms associated with the Data Model.

The National Cooperative Education Statistics System

The work of the National Forum on Education Statistics is a key aspect of the National Cooperative Education Statistics System. The Cooperative System was established to produce and maintain, with the cooperation of the states, comparable and uniform education information and data that are useful for policy making 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 focuses on 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 practices guides in data-related areas of interest to federal, state, and local education agencies. They are assisted in this work by NCES, but the content comes from the collective experience of the state and LEA task force members, who review all products iteratively throughout the development process. Documents prepared, reviewed, and approved by task force members undergo a formal public review. This public review consists of focus groups with representatives of the product's intended audience, review sessions at relevant regional or national conferences, or technical reviews by acknowledged experts in the field. In addition, all draft documents are posted on the Forum website prior to publication so that any interested individuals or organizations can provide feedback. After the task force oversees the integration of public review comments and reviews the document a final time, publications are examined by members of the Forum standing committee sponsoring the project. Finally, the entire Forum (approximately 120 members) reviews and formally votes to approve all documents prior to publication.

Contents

The Data Model Task Force ii
Acknowledgements iv
Foreword vi
Contents viii

Chapter 1: Introduction 1
How to Read This Guide 1
Background 2
The Products 3

Chapter 2: Education Data Model Overview 5
What is a Data Model? 5
The Benefits of Conceptual Education Data Modeling 7
Risks Associated With Conceptual Modeling 9
Rationale for Developing the Education Data Model: Version 1 (PK-12) 9
Uses and Users for the Education Data Model 10

Chapter 3: Navigating the Data Model 12

Chapter 4: Content and Structure of the Data Model 16
Core Data Model Concepts 16
Entity 16
Attribute 16
Common Attributes 16
Class 19
Taxonomy 20
Relationships 21
Concept Map 23
Data Model Content Structure 24
Content Domains 24
Data Needs Identification 25
Major Education Processes 26
Suggested Open Source Tools 28
Protégé 28
SWOOP 28

Chapter 5: Examples of How to Use the Data Model 29
Educator Vignette 29
Vendor Vignette 32
Researcher Vignette 33

Appendix A: Q&A 34

Appendix B: Glossary 36

Bibliography 37

Chapter 1: Introduction

The *Guide to the Education Data Model: Version 1 (PK-12)* explains the Education Data Model, how it was developed, and the benefits of the data modeling approach that was used. This document also describes practical ways of using the Education Data Model to address common challenges in the management and use of education data.

This guide is written for educators, information system /software vendors, researchers, and other potential users of the Education Data Model (occasionally referred to in this document as simply "the Data Model"). And, as it introduces concepts and approaches to structured data that may not be familiar to the average reader, it should also be helpful to anyone who may be involved in a project that uses the Data Model, but is not necessarily a data model expert.

How to Read This Guide

This guide provides an overview of the concepts and terminology of the Education Data Model. This introductory chapter discusses the objectives and suggested uses of the publication, provides background on how the Data Model was developed, and introduces additional products that support the model. Chapter 2 provides a general introduction to data models and discusses the rationale for developing the Data Model. Chapter 3 guides potential users through the navigation of the Education Data Model. Chapter 4 discusses the content and organization of the Data Model and provides some information on the process used to develop its contents. Finally, Chapter 5 includes a number of vignettes illustrating how different types of users might employ the Data Model to help them face various data challenges.

Neither this document nor the Education Data Model itself takes the place of a skilled data modeler, system architect, or designer. However, the Data Model is a resource that can be used in many ways and by many different types of education and technology professionals. As a reader, you might choose to read some parts of the document before others. Here are some suggested paths through this guide and beyond, to the Data Model itself.

- If you are interested in knowing how the Education Data Model could be used in your local education agency (LEA), and are just finding out about conceptual data models, read straight through chapters 1 and 2 in this document. Read the Core Data Model Concepts section in chapter 4. The sample uses of the Education Data Model in chapter 5 will also be of interest to you. Be sure to go to the Education Data Model website and study the use case scenarios that are mentioned in chapter 5, as well.

- If you are coming from a background in other data model domains, such as banking or healthcare, and want to familiarize yourself with the education data model domain, review the Common Attributes section in chapter 4; read chapter

5: Examples of How to use the Data Model; then go directly to the Education Data Model using the link provided in chapter 3.

- If you are an LEA or state education agency (SEA) policy maker, focus on chapters 2 and 5.

- If you are interested in localizing or extending the Data Model for use in your own organization, be sure to read Appendix A: Q&A, for technical details concerning the structure of the Data Model, and the Data Model Content Structure section in chapter 4, for information about how the Education Data Model content was produced.

Background

In 2005 the National Center for Education Statistics (NCES) commissioned a white paper to analyze existing data models used in the PK-12 education marketplace. A group of LEA, SEA, policy, and vendor experts concluded that the majority of data models in use are housed in vendor-specific data management systems, and that LEAs and SEAs needed a nonproprietary education data model to use as a blueprint to develop their own data models and systems, or to identify the requirements for purchased data management systems. As a result, the Technology Standing Committee of the National Forum on Educational Statistics convened a task force to develop a comprehensive PK-12 data model. The task force, made up primarily of LEA and SEA representatives, began its work by collecting currently used data frameworks and identifying what was needed for a generic education data model that could be used by LEAs, SEAs, federal policy makers, and education data management software vendors.

The newly approved task force examined existing data models, Forum publications, and other sources to seed the Education Data Model with content. Among other resources, the original sources used included the following:

- NCES Handbooks Online
- NCES Financial Accounting for Local and State School Systems
- NCES Forum Guide to Elementary/Secondary Virtual Education
- The Schools Interoperability Framework (SIF) Data Model
- Education Data Exchange Network (EDEN) submission file standards
- Input from Forum Data Model subgroups, including non-Forum experts
- Input from education software vendors

Some basic tenets of data model generation were identified from other sectors (healthcare, military, aviation, etc.) and were included in the initial data model design.

The content of the Education Data Model was developed by experts who worked in six task force subgroups, coming together in one-time external focus groups, and continuing their work through ad hoc communications. The experts included educators and administrators from schools, LEAs, and SEAs; policy makers; assessment and software developers; researchers; and software/systems vendors.

An early activity undertaken by the Education Data Model project included a national outreach to collect important education questions from education stakeholders. These questions were collected over a 6-month period and the results were categorized and summarized in a document—the *Forum Data Model Task Force Data Needs Survey Summary*—which can be accessed at: http://nces.ed.gov/forum/pdf/DataNeedsSummaryv1p0.pdf. The Schools Interoperability Framework Association (SIFA) was contracted to help the Task Force develop a data model that would satisfy the information needs embedded in these questions.

A qualitative analysis of the collected responses yielded nine broad categories of questions. These categories summarize the scope of the questions collected. These categories informed the more sophisticated frameworks that were used by the Forum Data Model Task Force to create focus for the subgroup work. The categories included questions about:

- Student progress over time
- Student performance and characteristics
- Program/plan performance
- Course of study effectiveness
- Teacher performance and characteristics
- School performance and characteristics
- LEA performance and characteristics
- State performance and characteristics
- Parents and the community

From these questions, the *entities* (items that need to be tracked in order to answer the questions) and the *attributes* (measures associated with the entities that are required to answer the questions) were identified.

Next, the entities and attributes were organized into a taxonomy or hierarchical organization of concepts based on similar and dissimilar characteristics. Finally, modeling the more complex relationships among entities produced the draft conceptual data model (See the *What is a Data Model?* section in Chapter 2).

The Products

The work of the Forum Data Model Task Force culminated in three products:

1. A comprehensive, community-developed Education Data Model that focuses on information needed to operate, improve, and report within the education marketplace.
2. An Education Data Model website, which includes supporting documents, use case vignettes, a web-based tool for browsing the Data Model, and links to open source tools that can be used for viewing, exploring, and applying the Data Model: http://nces.ed.gov/forum/datamodel. A second website is available to serve as the Education Data Model development website: http://nces.sifinfo.org/datamodel. This site is used for collecting community input

on the Data Model and will have Data Model tools that are under development and not yet approved and available on the official website.

3. This *Guide to the Education Data Model: Version 1 (PK-12).*

The Education Data Model: Version 1 (PK-12) is intended to be constantly thriving, expanding, and developing. The deliberate identification of this model as "Version 1" underscores that the Forum will continue to bring education stakeholders of all types to the conversation about the data needed at all administrative levels of education. The web tool and Forum activities should ensure that the Education Data Model will continue to grow in response to the needs of all education data users.

Chapter 2: Education Data Model Overview

What is a Data Model?

A data model is a structured description of how data are represented, organized, and accessed in an information system. A data model defines the data objects contained in an information system, the relationships between these data objects, and how the information is to be used. Data models are often represented as entity relationship diagrams. They support the development of information or data management systems, enable the exchange of data, enhance application maintainability, and can be imported in parts or whole to reduce the time and cost of developing new systems or redesigning existing ones.

Data modeling consists of a family of techniques used to describe the types of information important to an enterprise. It is a critical tool in providing the information that enterprises need to thrive and to comply with government regulations and industry standards. Data modeling is no longer limited to databases nor is it still viewed as just a tool for information technology (IT) staff. Rather, it is an increasingly expected tool for designing, visualizing, and communicating about education data, data systems, and analyses.

The enterprise of education requires continuous and wide-ranging information-based conversations to facilitate everything from teaching and learning processes to making federal policy decisions. Data modeling makes it possible to organize that information so that it can be used effectively and efficiently.

The American National Standards Institute describes three kinds of data models:[1]

1. A *conceptual* data model describes a content domain, such as education information, that consists of significant concepts and the relationships among the concepts in the domain. The meaning of a concept is communicated not only by text descriptions, but also by the concept's context in the web of relationships with other concepts. The relationships between pairs of concepts can be expressed in a formal logic language that is independent of any data manipulation technology or implementation environment.

2. A *logical* data model describes a subset of a content domain, such as student information or special education program information, in terms of a particular data manipulation technology. For example, if the data manipulation technology is a relational database, then the content is expressed in terms of tables and columns. If the data manipulation technology is XML[2], then the content is

[1] American National Standards Institute. 1975. "ANSI/X3/SPARC Study Group on Data Base Management Systems; Interim Report". FDT (Bulletin of ACM SIGMOD) 7:2.
[2] Extensible markup language (XML) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere.

expressed in terms of XML objects and XML elements. Logical models do not contain details of the implementation environment.

3. A *physical* data model describes a logical model in terms of the implementation environment. The implementation environment includes the specific product being used to manage the data and the physical means by which data are stored, such as servers and disk drives, lookup tables, management tables, etc.

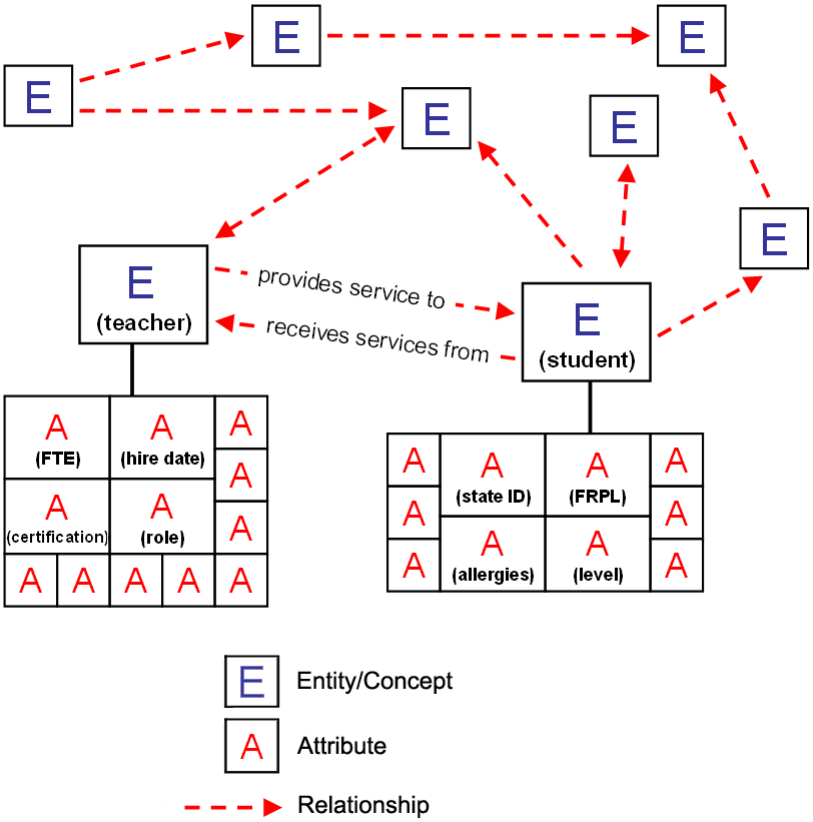
Database engineers and designers use physical and logical models to design and implement their systems. Conceptual models are convenient tools to help end users and other stakeholders discuss, understand, and participate in the design of systems, such as education data systems or education delivery systems. The Education Data Model is a conceptual data model.

The Education Data Model: Version 1 (PK-12) depicts a large portion of the information that should be contained in education software, information systems, and data warehouses. As a high-level conceptual model, the Education Data Model outlines the meaning of concepts and the relationships among concepts without imposing any particular logical or physical implementation approaches. (See chapter 4, especially the *Concept Map* section for further information about the model's content.)

The information in this conceptual model takes into account the processes associated with teaching, learning, and the business operations of education organizations. The Data Model focuses on granular information at the school and LEA levels, rather than upon aggregate statistics or indicators for accountability; for example, it includes "student" as an entity but not "total number of students." However, the Data Model includes information that is necessary to produce aggregate and other types of statistics. In brief, the Education Data Model is a catalogue of the data used in PK-12 education and a description of the relationships among those data.

Figure 1 depicts the concepts, relationships, and attributes of the Education Data Model. In the Data Model and the figure below, concepts are called entities (constructs that need to be tracked, measured, and described by software systems in order to support education processes). These entities have relationships to one another, reflecting the associations that exist between them in the real world. In addition, entities have attributes (information associated with an entity that can be measured, classified, or described) (see the *Core Data Model Concepts* section of Chapter 4). These entities, relationships, and attributes provide the basis for answering important education questions and designing complete information systems.

Figure 1. Concepts, Relationships and Attributes of the Education Data Model



For instance, in the Education Data Model, teachers and students are entities. These entities are related because teachers provide services to students, and conversely, because students receive services from teachers. Teachers have attributes such as full-time equivalency, certification, hire date, role and content knowledge. Students have attributes including state ID, name, free and reduced price lunch eligibility, allergies, and course completion records. Later in this guide, the structure of the data model will be discussed in greater detail, showing how the entities are organized into a broad taxonomy of classes and subclasses (see chapter 4 for further information).

The Benefits of Conceptual Education Data Modeling

Comprehensive data models can be invaluable in helping education respond to changing business conditions. When used appropriately, conceptual data models can lead to more accurate data, which in turn support more effective use of data contained in business-intelligence and business-analytic tools. Education stakeholders at the LEA and SEA levels who understand basic data-modeling concepts and who work with effective data models are likely to be especially productive. Conversely, information workers who have to struggle with ineffective models or who are not literate in data models face a couple of unhappy choices. They can rely solely on IT counterparts or vendor-specific solutions to assist them with data analyses, and may not get exactly the information they need; they can skip the analyses altogether; or, potentially the most dangerous solution, they can

base analyses on inaccurate or incomplete data. Some possible benefits of a well-developed data model are described in the following subsections.[3]

Process Design: The very process of collaboratively designing a data model has substantial benefits. Designing a local data model can bring together all stakeholders and give them a better understanding of their own data needs and those of their coworkers. The process can also help stakeholders better understand the needs and value of reporting authorities, and drive data decisions and uses with the input of stakeholders' diverse education expertise rather than depend on IT staff to make all determinations.

Communication with Stakeholders: Education shares with other endeavors the tendency to operate in various silos, based on the belief of each silo's owner that it represents unique needs and functioning. Data models, when used strategically, can stimulate communication between educators who are the silo owners and with various external stakeholders requesting program information – policy makers, the media, local and state agencies outside of education, and parents. When stakeholders share a common and comprehensive picture of all the data needed in the education enterprise, individuals in the various stakeholder roles can more readily and accurately communicate their needs.

Employee Training: Comprehensive data models can provide significant professional development opportunities for data users. Staff who are vested owners of a developed data model become better informed in all aspects of the education enterprise and the various dependencies on information. In local systems development, each employee adds expertise to the model to ensure completeness and clearly articulate the data needs to be met. An additional benefit is that an understanding of the need for quality data becomes embedded in the culture of the entire organization.

IT Benefits: Accurate data models lead to more productive and usable data systems at all levels of the organization. Input from the eventual consumers of data systems enables IT professionals, including vendors, to design and implement systems that accurately depict information needed to support daily operations, teaching and learning, and various data reporting demands: information systems that work.

Reporting and Operational "Alignment": Well-developed data models can have dramatic effects outside the daily operations in education. Data models can be used to align data collected and maintained in local systems with data needed for state and federal reporting entities such as the accountability reporting for No Child Left Behind. When data models in local systems vary greatly from the data model of the mandating entity, the reporting process becomes inefficient and increases the burden on the reporting entities, leading to the complaint that schools are asked to report what is basically the same information over and over again. This extra burden ultimately affects data quality and can result in misinformation.

---

[3] Maguire, Joe. 2008. *Data Modeling: A Necessary and Rewarding Aspect of Data Management*. Midvale, UT: Burton Group.

Risks Associated With Conceptual Modeling

One of the most common stumbling blocks in developing and using data models is viewing them first and foremost as IT projects. Successful data model projects first focus on the business and use case needs of eventual data consumers. A strength of the conceptual Education Data Model is that it is technology-independent and was developed by individuals in various stakeholder roles. Before technology becomes the focus of a data modeling project, the considerations described in the following subsections should be taken into account.[4]

Product Scope: If the scope of a data model project is not carefully thought out before other work begins, the project may end up with the wrong product – or no product at all. During the development of the Education Data Model Version 1: PK-12, the task force carefully defined the scope of the two year project. As a first step, the task force determined that the Data Model would focus only on PK-12 data—more specifically, on the data interactions and definitions supporting the student, teacher, and course triad. Although it is expected that the Data Model will continue to develop over time, defining the initial scope allowed the task force to complete Version 1.

Incompleteness: As with data systems, a data model is never really complete. Instead, it should be viewed as dynamic information that provides ongoing opportunities for input and for addressing evolving information needs. In this project, the task force made every effort to identify the most basic and common data needs at the school and district levels, and left itself open to add future directions.

Generalization: In effective data model development, it is important that specific bits of data link via specific relationships to other bits of data. In building conceptual data models, it is easier to make general linkages between categories of entities than to look for the specific relationships that provide the most detailed information for users. Attaining this kind of specificity is difficult, but the standardization that results from using this kind of granular information is beneficial. The Education Data Model is built on specificity rather than on general linkages.

Rationale for Developing the Education Data Model: Version 1 (PK-12)

The Education Data Model is a comprehensive, localized, conceptual model that provides a generic blueprint for schools and districts. This blueprint enables schools to evaluate and improve instructional tools, communicate those needs to their umbrella agency or directly to vendors, enhance the movement of student information from one LEA to another, and in the end, have better tools to inform instruction. Using a standard Education Data Model as a starting point contributes to a comprehensive understanding of the need for data, how data are used, and the questions that can be answered with the data. For instance, the Data Model helps to answer questions such as the following:

- What data do schools, LEAs, and states need to collect and manage at the local level to meet the information needs of students, staff, and other stakeholders?

[4] Maguire, Joe. 2008. *Data Modeling: A Necessary and Rewarding Aspect of Data Management*. Midvale, UT: Burton Group.

- What data do they need to effectively manage education organizations in order to increase success in teaching, learning, and school leadership?
- What data do they need to efficiently manage and run an education organization from a fiscal and administrative perspective?

Schools and LEAs that have data models often rely on proprietary models developed by vendors and implemented in vendors' software applications. With two out of three (69.6 percent)[5] districts in the United States enrolling fewer than 2,500 students, many LEAs either cannot afford proprietary data solutions, or cannot afford to tailor purchased models to the needs of their education stakeholders. States may lack the financial or technical means to develop comprehensive data models for their LEAs.

Until now, there has not been a comprehensive, nonproprietary, generic education data model for use by schools, LEAs, and states to design or guide the selection of systems for instructional delivery, decision support, operations, reporting, and data warehousing. The Data Model serves as a tool to convene relevant stakeholders around the evaluation of data management processes at the school and LEA levels. The model works as a tool to help agencies through the process of identifying requirements, and can then be used to communicate these requirements to vendors (or internal staff in the case of a build solution). And, the model provides a template against which agencies can assess the adequacy of proprietary systems to meet their specific needs.

Uses and Users for the Education Data Model

The Data Model is intended to serve multiple purposes for multiple audiences. Some potential users who should find its benefits relevant to their roles and responsibilities include:

- **Local educators, administrators, and technology directors:** To help spur data management discussions and define needed data elements when designing, choosing, or augmenting education software or data systems. For example, the Data Model can help
  - schools evaluate their own local data input and management from a multiple stakeholder perspective (clerical, administrative, educator, etc.);
  - policy makers become more aware of the range of available data and form appropriate policy questions; and
  - request for proposals writers determine more accurately and efficiently the range of information needed to meet the requirements of a desired system.

- **Software developers:** To ensure that software products contain standard information needed by all PK-12 organizations. For example, the Data Model can help

[5] Table 84: Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2008). Digest of Education Statistics 2007 (NCES 2008-022). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

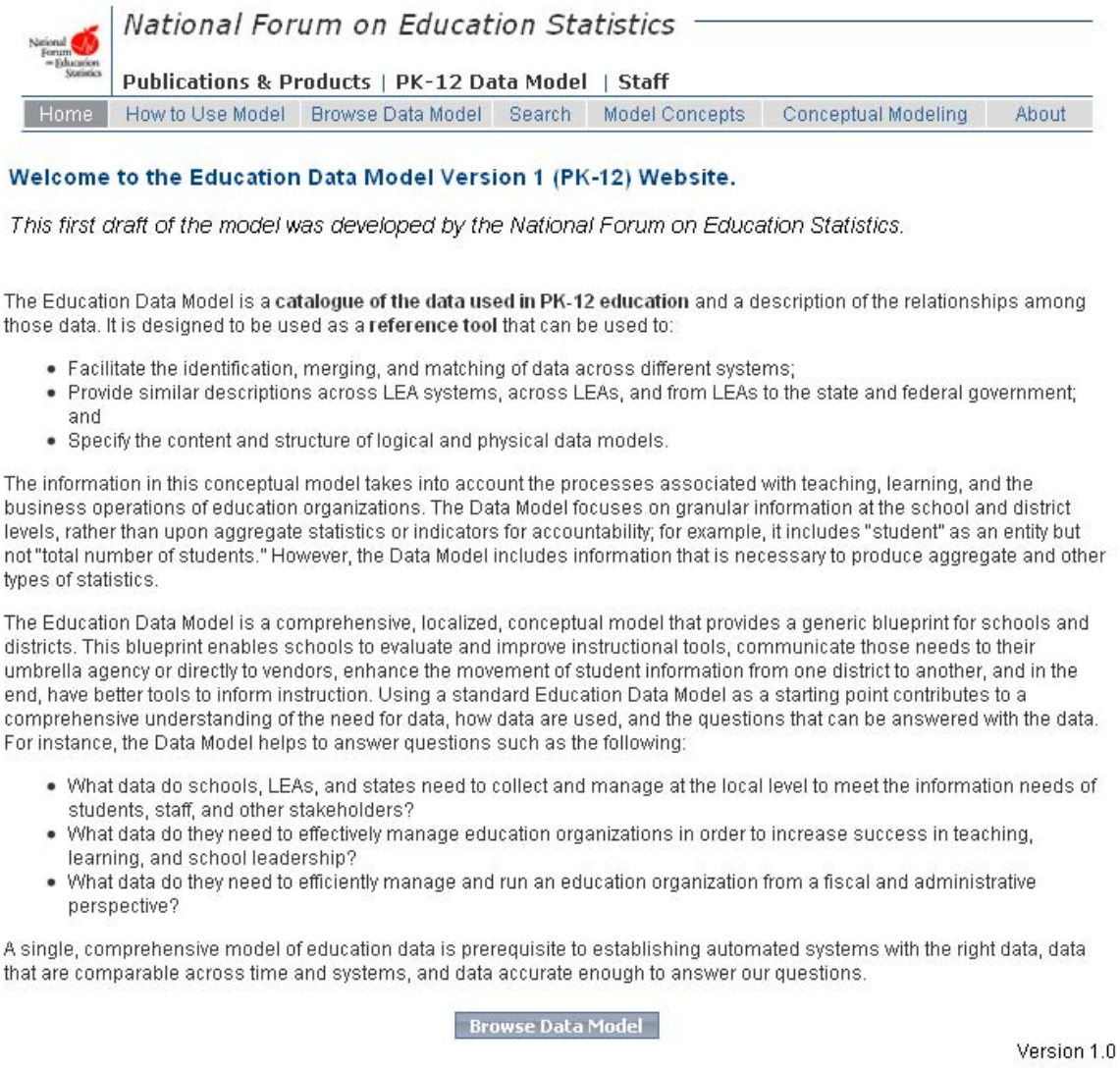
  - vendors develop systems that more accurately support management, analysis, and reporting; and
  - responders to requests for proposals understand the appropriate range of information for the systems and services to be provided.

- **Researchers:** To identify what information is currently available in schools and what might be available in the near future. For example, the Data Model can help
  - researchers explore the composition of data sets for analysis and communicate their data needs.

- **All constituents:** To communicate about the education-related information that should be collected now and in the future, in order to foster research and practice. For example, the Data Model can help LEA and SEA administrators better identify analytical needs and the necessary data for continuous improvement planning.

Chapter 3: Navigating the Data Model

This chapter describes each menu item in the navigation box on the home page of the Education Data Model website. The Data Model and its accompanying documentation can be accessed at http://nces.ed.gov/forum/datamodel.

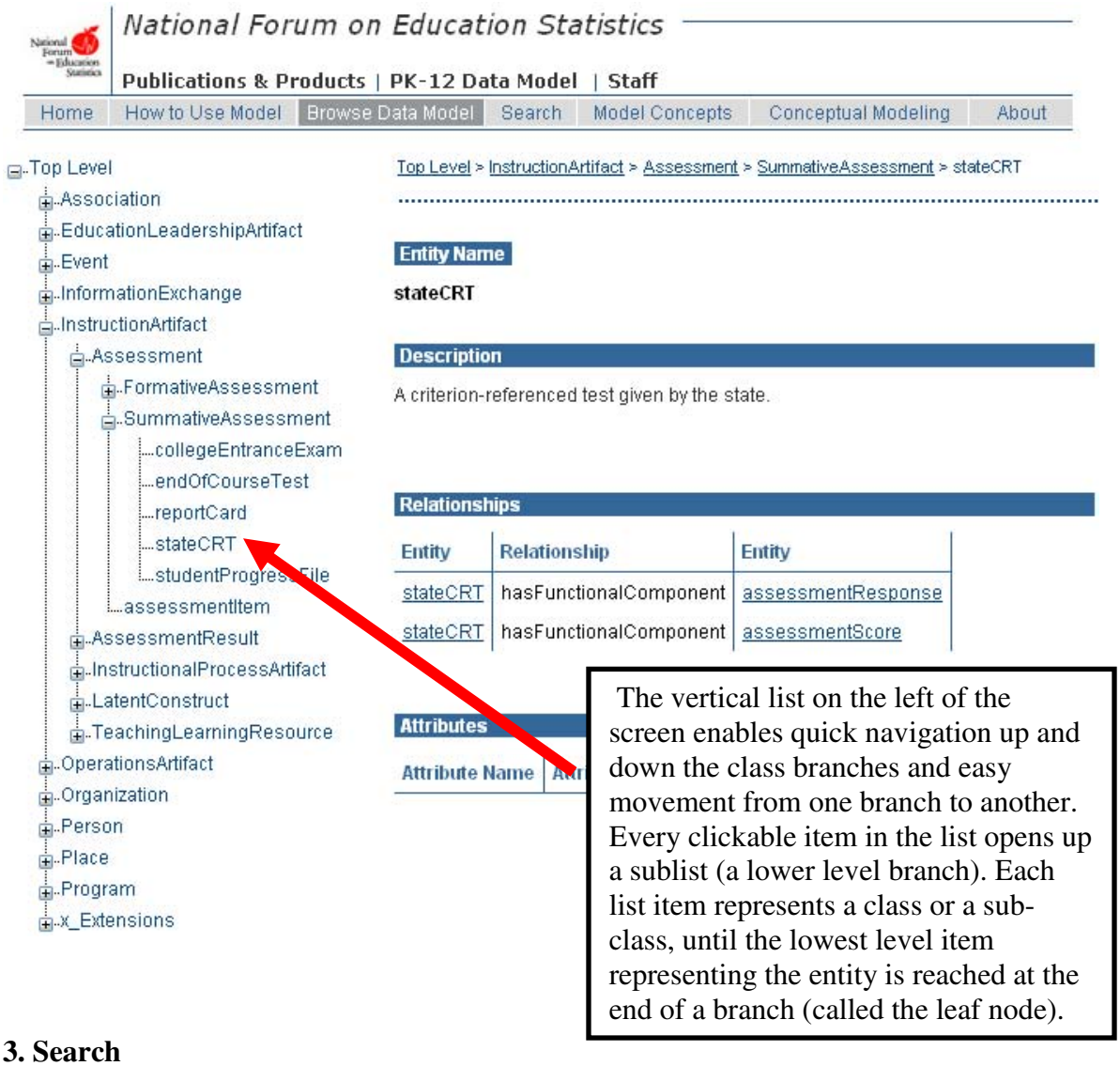


1. How to Use Model

The first link on the home page navigator bar provides information that explains how to make use of the content contained in the Data Model. Links to other pages in the website are provided along with a reference to this guide. One of these links brings users to the How to Browse Data Model page, where they can find printable step-by-step explanations for using all the features of the Data Model browser.

2. Browse Data Model

The second link navigates to the web-based Data Model browser tool. The browser enables users to view the Data Model and perform searches without having to install additional software. The browser enables the user to view and browse the Data Model taxonomy. Each class in the taxonomy has a screen that lists all of its sub-classes, which are the entities contained in the class. Each entity in the taxonomy has a screen that lists its attributes and relationships to other entities.



The vertical list on the left of the screen enables quick navigation up and down the class branches and easy movement from one branch to another. Every clickable item in the list opens up a sublist (a lower level branch). Each list item represents a class or a sub-class, until the lowest level item representing the entity is reached at the end of a branch (called the leaf node).

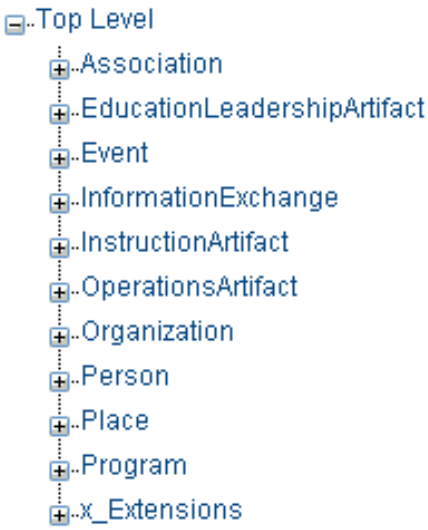
3. Search

The third link brings users to a Search tool, which enables searching the Data Model for

- Entity Name
- Entity Description
- Attribute Name
- Attribute Description
- Class Name
- Class Description
- Relationship Name

4. Model Concepts

The fourth link describes core concepts associated with the Data Model and the ten top-level classes of the Data Model taxonomy. The list on the left of the Data Model Browser screen enables the user to "drill down" into the sub-classes that are below each top-level class. The ten top-level classes are described below.



Association: Functional associations among individuals, organizations, events, and programs. Examples of associations include student enrollment in a school, a contact person for an organization, and a sponsor for an event.

Education Leadership Artifact: Artifacts related to the leadership of teaching and learning processes. Examples of education leadership artifacts include school improvement plans, performance criteria, and lesson plans.

Event: Any event within the education organization. Examples of events include billing transactions, calendar events, and school terms or meetings.

Information Exchange: Information being communicated. Examples include an email, a press release, and data collection from surveys and reports.

Instruction Artifact: Any piece of information related to a student's learning or the general learning process. Examples include entities related to teaching and learning, such as learning resources, assessment scores, learning standards, and observations.

Operations Artifact: Information related to the administrative or operational side of the organization. Examples include compensation for employees, capital assets, expenditures, and teaching credentials.

Organization: Groups of resources and people that exist to fulfill a particular role or achieve a particular goal. Examples include education institutions, LEAs, SEAs, and school organizations such as a parent organization.

Person: Individuals within the education setting. Examples include a student, a teacher, and a parent volunteer.

Place: The locality or area within the education environment. Examples include the school district boundaries, a bus position, and an attendance zone.

Program: A plan of activities and procedures to accomplish a set of objectives. Examples include a federal program, a staff development program, and a course sequence.

x_Extensions: This class is not a regular part of the Data Model. It is a category that contains extensions to the Data Model. For example, the category could hold Education Data Exchange Network (EDEN) input file definitions or fact table definitions for a data warehouse. This category allows concepts that are not allowed in the Data Model, such as aggregate statistics or report definitions, to be linked to entities and relationships in the Data Model.

5. Conceptual Modeling

The fifth link enables users to download the Data Model in an OWL zip file. The downloaded file can then be used with various data model tools, including two suggested open source tools (see the *Suggested Open Source Tools* section in Chapter 4), to manipulate and localize the Data Model.

OWL stands for Web Ontology Language (first two letters reversed). OWL is an XML language for expressing relationships and is a standard published by the W3C web standards group. The OWL language represents both a taxonomy of entities, as well as multiple types of simultaneous relationships among the entities. It is also designed to be compatible with the World Wide Web (http://www.w3.org/2004/OWL/).

6. About

The sixth link provides information outlined in this Forum publication concerning the rationale, methodologies, and contributors in the development of the Education Data Model: Version 1 (PK-12).

Chapter 4: Content and Structure of the Data Model

This chapter describes the contents and organization of the Data Model. It begins with a discussion of the core concepts and terms related to the Data Model. Then, the chapter illustrates the Data Model's content structure. Two open source tools available for viewing and manipulating the Data Model are presented at the close of the chapter,

Core Data Model Concepts

This section explains important terms and concepts related to the Data Model: entity, attribute, common attributes, class, taxonomy, relationship, and concept map.

Entity

An entity is the basic building block of the Education Data Model. Entities are the constructs that need to be tracked, measured, and described by software systems in order to support education processes. Entities can be:

- Persons such as students, parents, or staff members
- Capital assets such as schools, school buses, or buildings
- Events such as test administrations, class meetings, or discipline incidents
- Tools such as books, networks, computers, lessons, or assessments
- Concepts such as perceptions or skills

Attribute

An attribute is information about an entity that you can measure, classify, or describe. An attribute is not a calculation or statistic, and it generally does not contain counts.

Attributes in the Data Model are not generic measures or data elements that can connect to a number of entities. In the Data Model, each attribute is unique to the entity that it measures. The following are examples of attributes:

- A measurement, current state, or a trait of an entity
- A person's name, phone number, or IM address
- An education institution's name, location, or size
- The date of a test administration, the value of an assessment score
- A lesson topic, grade level, or student's learning style
- A teacher's perception of self-efficacy with respect to classroom management

Common Attributes

Attributes are usually specific to each entity; however, there is a small set of attributes that apply to multiple entities[6]. These attributes are referred to as common attributes and are arranged into a taxonomy. This taxonomy of common attributes should not be confused with the Data Model taxonomy of entities.

[6] Definitions of the common attributes can be found the NCES Handbooks Online at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008805.

A limited set of the common attributes used in the Education Data Model is listed below. The list is provided to establish consistency and uniformity in the naming and use of these attributes. These common attributes involve names, location information, and identifiers and are expected to change as the Data Model continues to be updated.

Items lower in the hierarchy are more specific than those above. Items at any level in the hierarchy are directly or indirectly associated with at least one entity.

Person Name

- Alias
- First Name
- Former Legal Name
- Generation Code/Suffix
- Last/Surname
- Last/Surname at Birth
- Middle Name
- Nickname

Locus

- Location
- Physical Location
    - Physical Address
    - Latitude/Longitude
    - GPS Coordinates
- Virtual Location
    - IP Address
    - URI
    - Postal Address
- Connection ID
    - E-mail Address
    - Phone Number
    - Screen Name
    - IM Address
- Schedule
    - Day/Date/Time
    - Periodicity
    - Length

Person Characteristic

- Demographic
    - Race and Ethnicity
    - Sex/Gender
- Role
- Physical Characteristic
    - Weight
    - Picture/Likeness

- Status

Document Metadata

- IEEE LOM
- Dublin Core
- SIF
- RDF

Program Evaluation

- Resources Consumed
    - Financial Resources
    - Time Resources
    - Human Resources
    - Curricular Resources
    - Instructional Materials
        - New Technology Resources Used
        - Equipment Used
        - Supplies Used
    - Facilities Resources
        - Number of Buildings
        - Square Feet
- Measurable Goals and Outcomes
    - Academic Goals
        - Academic Achievement
        - Skills Acquisition
        - Skill Certification
    - Non-academic Goals
        - Health Effects
            - Physical
            - Emotional
            - Developmental
        - Social Effects
            - Community Service
        - Participation Effects
            - Retention
            - Enrollment
            - Completion
    - Goal-Based Outcomes
- Target/Served Population
    - Size
    - Demographics
    - Program Baselines
    - Method of Identification
    - Population Characteristics
- Methodology
    - Measurement

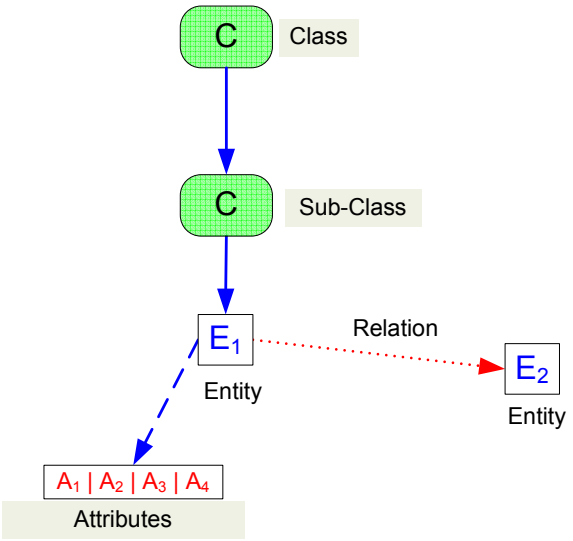
    - Analysis
- Characteristics
    - Program Availability
        - Geographic
        - Calendar
        - Length of Program
        - Periodicity of Service
        - Length of Services
    - NCES Program Type
    - Delivery Methodology
        - Constructivism
        - Inquiry Based
        - Directed Instruction
        - Virtual
        - Groupings

Class

A class is "A set, collection, group, or configuration containing members regarded as having certain traits in common; a kind or category."[7] To distinguish between entities and classes, the Data Model begins the names of classes with a capital letter and entity names with a lower-case letter. For example a "Person" is a class while a "staffMember" is an entity.

Figure 2 illustrates the relationships between classes, entities, and attributes. *Entities*, which identify what needs to be tracked by education systems, are organized into *classes* and *sub-classes*. Each entity may have *attributes*, which represent the measures that are used to track the entity. In the Data Model, entities may relate to other entities.

Figure 2. Core Concepts of the Education Data Model

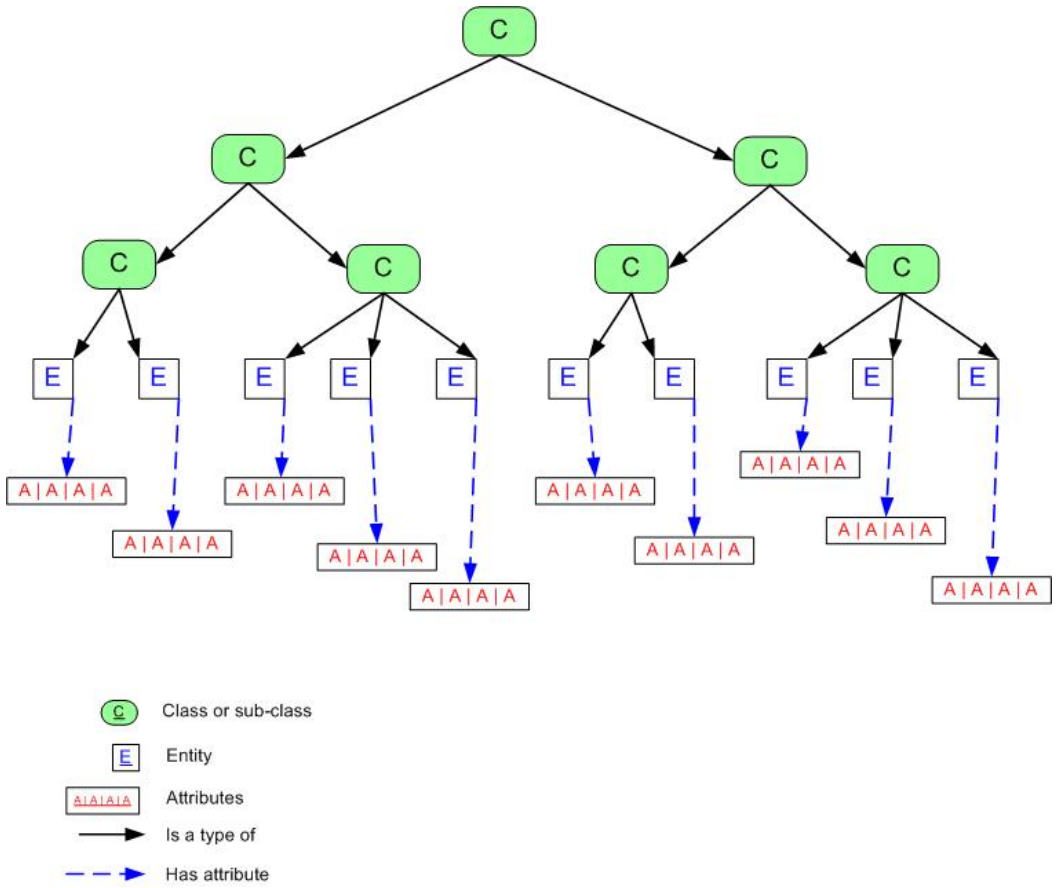


[7] *The American Heritage® Dictionary of the English Language,* (2004) Fourth Edition. Houghton Mifflin Company. http://dictionary.reference.com/browse/class (accessed: July 2008).

Taxonomy

The entities, classes, and attributes in the Data Model have been organized into a taxonomy. As illustrated in figure 3, entities (marked with E) organize into classes and sub-classes (marked with C). Each entity has its own set of attributes (marked with A). Entities group together based upon common characteristics.

Figure 3. Taxonomy of Classes, Entities, and Attributes



Taxonomies arrange items into categories based on like characteristics. For example, lesson plan and unit plan are both types of academic plans in the Data Model, but lesson plan and unit plan differ based upon the scope and purpose of the plan. The "is a type of" organization scheme ensures, with few exceptions, that each entity has one, and only one, place in the arrangement. For example, in the Data Model

> a portfolio *is a type of* formative assessment *is a type of* assessment *is a type of* instruction artifact,

just as in the taxonomy of living things,

> a Homo sapiens *is a type of* hominid *is a type of* primate *is a type of* mammal *is a type of* chordate *is a type of* animal.

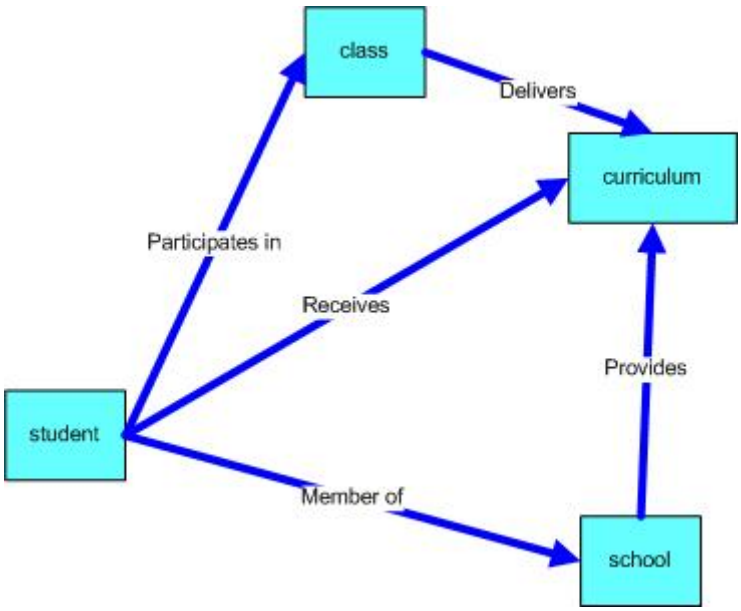
There are exceptions and hard-to-classify cases, similar to the duck-billed platypus in the taxonomy of living things, which, as a mammal that lays eggs, defies a clean classification.

The Data Model taxonomy is similar in form and function to other well-known taxonomies, such as the Linnaean taxonomy of living things, and the historically used classification systems, such as the Dewey Decimal System.[8] As the Data Model user becomes familiar with the structure of the taxonomy, locating a particular entity becomes easy. In addition, the tools provided for using the Data Model have search features that facilitate locating items in the taxonomy.

Relationships

In addition to the taxonomy structure, which is based on entity characteristics, the Data Model contains natural relationships among the entities. For example, in the Data Model taxonomy, the "student" entity is not close to the "class" entity, but the Data Model stores and reflects the relationships between them. Figure 4 shows examples of the types of relationships contained in the Data Model.

Figure 4. Types of Relationships in the Education Data Model



[8] Two other projects that will have influence on the Education Data Model are the Suggested Upper Merged Ontology (better known as SUMO) and the Universal Data Element Framework (UDEF). SUMO is an attempt to create a high-level model that applies to all domains of interest. SUMO is expressed in a formal logic language. At some point in the future, it would be useful to map the Education Data Model to SUMO. The UDEF is similar to the Education Data Model in that it attempts to model a particular domain (enterprise information systems). The Education Data Model has independently developed many classes that are similar to classes in the UDEF. In the future the Data Model may continue to be informed by the UDEF.

The relationship descriptors include verbs or short verb phrases that connect the subject with the object. For example, "student" (subject) "receives services from" (relation) "teacher" (object). This information within the Data Model allows for intelligent searching and for creation of subparts of a model. As the Data Model grows, so too will the types of relationships captured among its entities. Below is a list of relationships that is expected to change as the Data Model continues to be updated.

- **isDirectProviderOf**
  Directly provides goods or services.

- **isEvaluationResultOf**
  An evaluation of a student or other learner's performance.

- **hasDelimiter**
  One entity is modified or limited by another. For example, a calendar date can delimit a count of students.

- **determines**
  Determines in part or in whole the content, structure or value the object entity.

- **hasAssociated**
  The object entity must have the associated subject entity in order to exist. This is similar to a foreign key relationship. For example, a disruptive event (subject) has an associated victim (object).

- **hasCausalRelationship**
  A strong relationship in which one entity causes a change in another entity.

- **isAValueOf**
  Used to enumerate the possible values, or states of an entity.

- **receivesServicesFrom**
  Subject receives services from the object.

- **isPartiallyDefinedBy**
  The subject is defined in part by the object. Or, the subject is a narrower concept than the object.

- **hasFunctionalComponent**
  Reflects the construction of an entity through functional components represented by other entities.

- **participatesIn**
  A person-type entity participates in an activity-type entity.

- **isIndirectProviderOf**
  Indirectly provides goods or services.

- **isACountOf**
  A non-duplicated count

- **isDerivedFrom**
  One entity is a derivation of another. This means that some or all of the important features of an entity is also in its sub-entities. This is different from functional components.

- **providesServicesTo**
  Subject receives services from the object.

- **isOrganizationalComponentOf**
  This relation is used to indicate an organizational structure of non-person entities such as schools, districts, etc.

- **isFunctionalComponentOf**
  This relation indicates that subject entity makes up, in part or in whole, the function of the object entity.

- **isAlignedWith**
  Used mainly in Teaching and Learning entities. One entity is constructed to align in meaning or function with another.

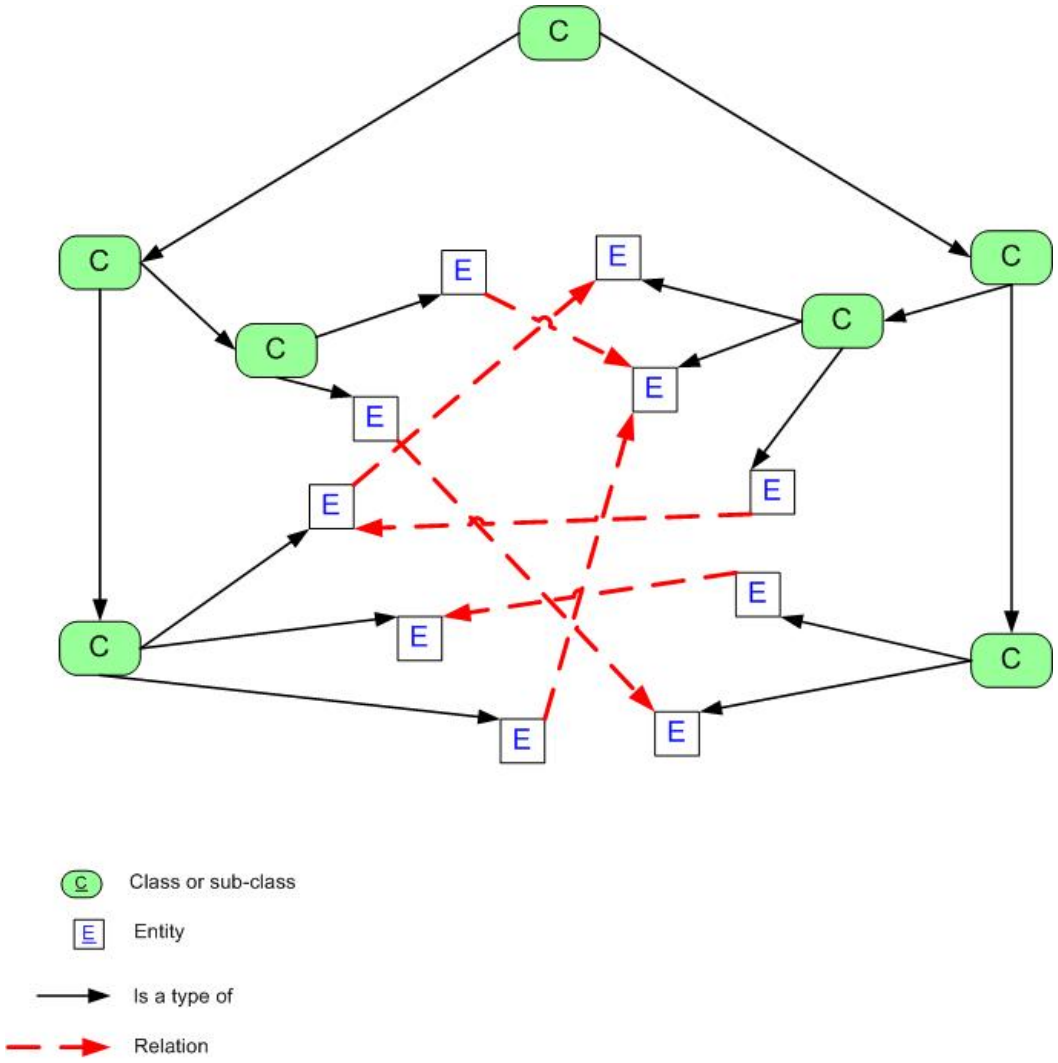
- **isASynonymOf**
  Similar in meaning. This could indicate an exact or inexact similarity.

Concept Map

The Concept Map describes the structure of the Education Data Model. It represents a logical and finite set of relationships among classes, sub-classes, and entities, thereby striving to depict the entire domain of education information. It adds multiple simultaneous relationships among the entities to the taxonomy. The relationships among entities are designed to be mutually exclusive but may sometimes overlap in meaning or usage.

To feature the relationships among the entities presented previously in figure 3, figure 5 turns the taxonomy inward.

Figure 5. Taxonomy Featuring Relationships Among Entities



Data Model Content Structure

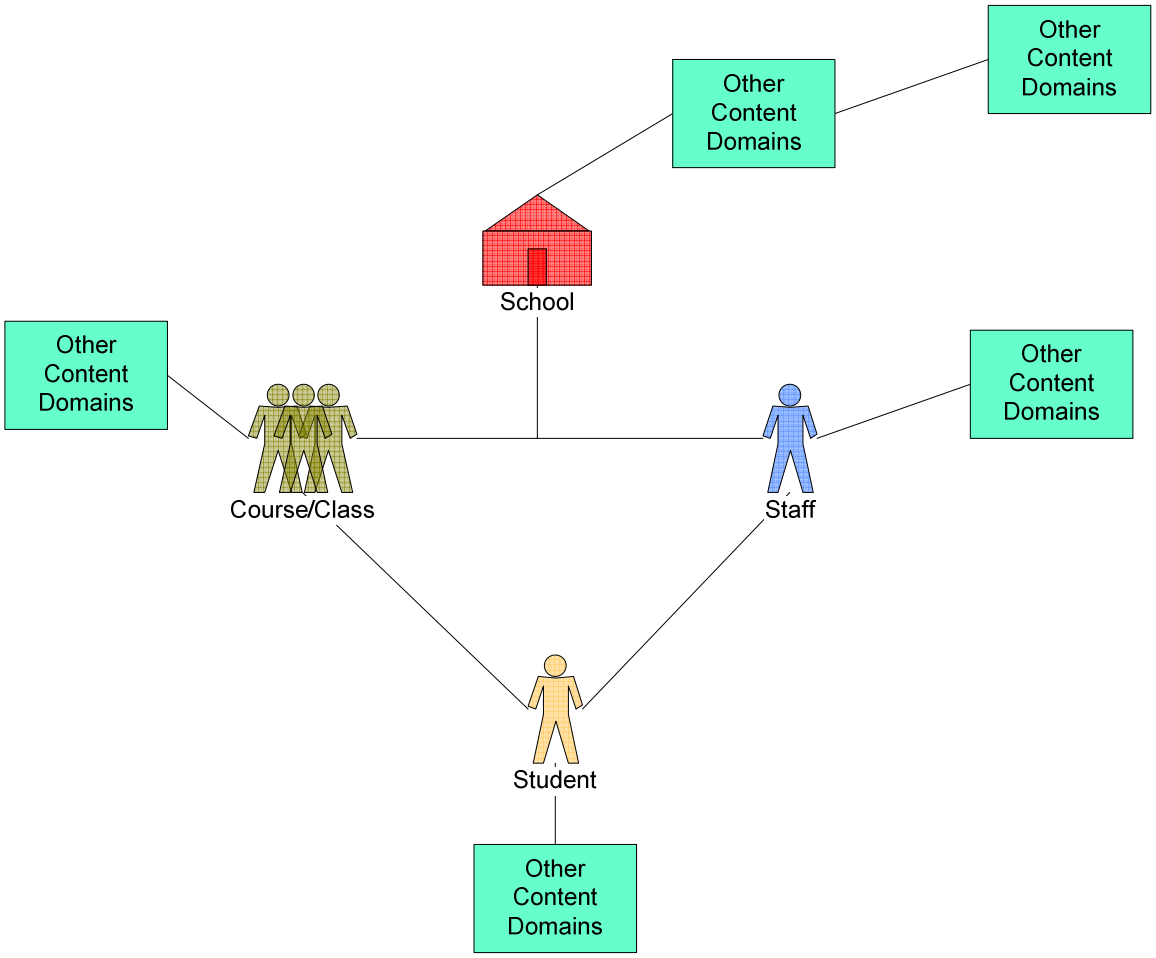
This section offers insight into the Data Model development process by describing the major concepts around which the Data Model was created. Two core content domains, teaching and learning, provided the foundation for Data Model development. A process of data needs identification was used to identify the information necessary to answer educators' questions. Several major education processes, or process perspectives, further informed the development of the Data Model's contents and organization.

Content Domains

The Education Data Model uses the lowest level of data granularity within PK-12, usually the individual, school, and LEA levels, where data originate and have the greatest impact on the teaching and learning process. Figure 6 illustrates the core domains of the Data Model "spine": teaching and learning. Especially important in this depiction is the

triad of student, staff (teacher), and course/class. The interactions among the three members of the triad formed the core processes of interest in the development of the Data Model. Other domains have been and continue to be developed around this core interaction, such as programs, activities, assessment, transportation, facilities management, professional development, and accountability.

Figure 6. Core Content Domains of the Education Data Model

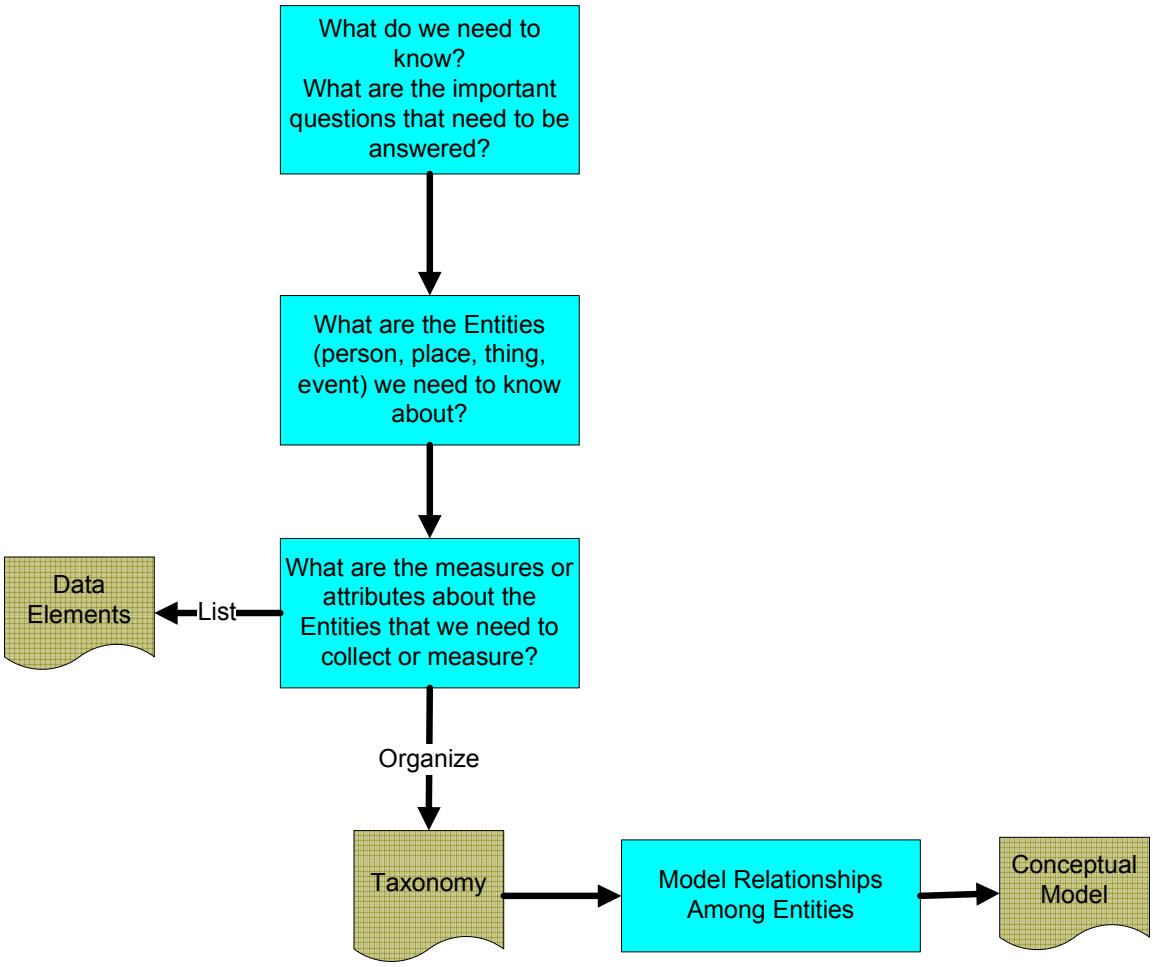


Data Needs Identification

The contents of the Data Model were generated by first asking "What do we (educators) need to know?" and "What questions do we need to answer?" The next step was to identify which people, things, and constructs ("entities" in data modeling language) we need to know about. The final step in generating the elements of the data model was to identify the facts that we need to know about the people, things, and constructs in order to answer our questions. The resulting data elements were organized into a hierarchical taxonomy.

Once the data elements were determined and organized, the relationships among them were identified and recorded. Figure 7 shows the described process.

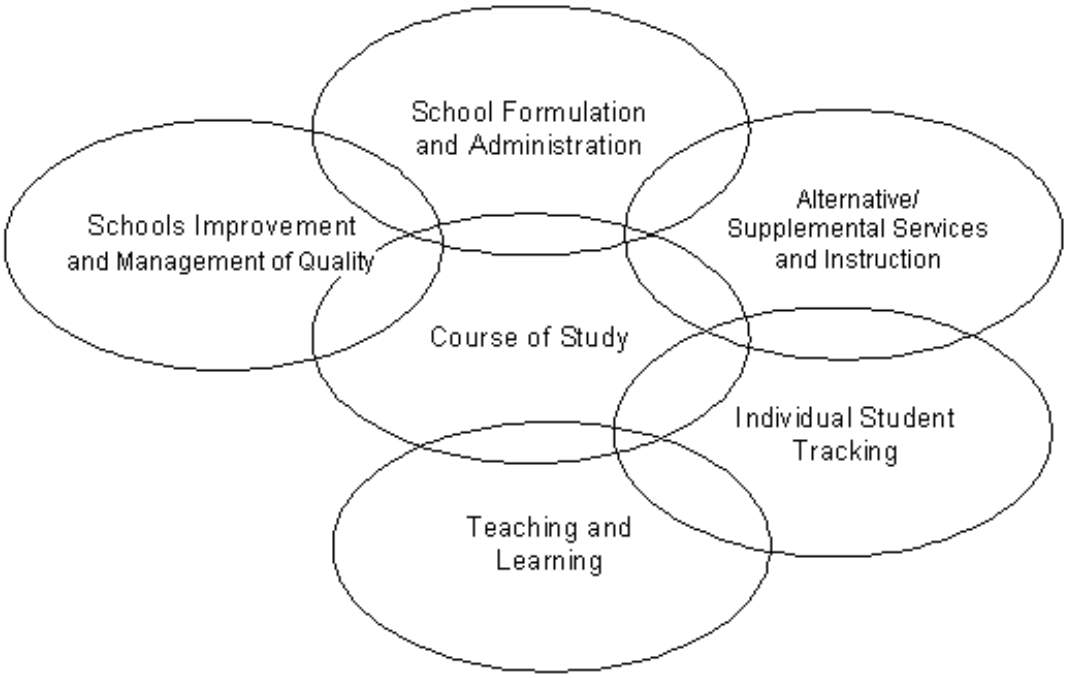
Figure 7. Data Needs Identification Process



Major Education Processes

The contents of the Data Model were also generated by examining six major education processes. These processes represent different perspectives from which items and relationships in the model can be generated. Together, the following six process perspectives make up the education enterprise. The overlap among these major processes, depicted in figure 8, defines the ways in which they interact to support the overall education enterprise.

Figure 8. Major Education Processes in the Education Data Model



School Formulation and Administration: Processes that address setting up and maintaining the physical, virtual, and community structures that make up a school.

Course of Study: Processes related to constructing and managing the academic environment around which learning takes place.

Alternative/Supplemental Services and Instruction: Processes that address programs, services, and instruction that supplement, or are alternatives to, the standard course of study.

Teaching and Learning: Processes directly related to teaching and learning.

Schools Improvement and Management of Quality: Processes related to the improvement of schools and strategic planning.

Individual Student Tracking: Processes that manage student information related to student status, student characteristics, evaluation of education programs, and accountability for student success.

Suggested Open Source Tools

More advanced users may want to display the information contained in the Data Model in ways tailored to their particular needs or interests. Some suggestions include two "open source" tools listed below, Protégé and SWOOP, which allow the user to accomplish tasks such as using the model to generate:

- Lists of entities and attributes
- Entity detail reports
- Conceptual maps
- Filtered conceptual maps, i.e., only portions of the conceptual map shown for ease of understanding
- Entity taxonomy
- Relationship diagrams

Protégé

Protégé is one of the most widely used tools to create relationships in OWL. It was developed by Stanford Medical Informatics: http://smi.stanford.edu/. The Protégé 3.4 full version software

- is an open source OWL and ontology editing tool,
- runs on multiple operating systems, and
- can be accessed at http://protege.stanford.edu/download/registered.html (be sure to install the OWL plug-in).

SWOOP

SWOOP is an OWL relationship browser and editor. It has an easy browser-like interface. The software was originally produced by the MINDSWAP group, University of Maryland, College Park. SWOOP software

- is an open source project with multiple individual contributors, and
- can be accessed at http://www.mindswap.org/2004/SWOOP/.

Chapter 5: Examples of How to Use the Data Model

The goal of the Forum Data Model Task Force was to produce a conceptual data model that is both comprehensive and useful. Most K-12 education agencies in the United States utilize data models embedded in the various purchased software applications at the school, LEA, or state level. The Education Data Model provides information to make all purchasers of such software and systems informed consumers, and to facilitate productive conversations between users and software developers in discussing accurate data system features and functionalities. The Data Model is a blueprint for identifying information needs and establishing expectations between educators, administrators, and software developers. This chapter discusses how educators, vendors, and researchers might use the tools described in Chapter 3 to utilize the Data Model.

Following are three vignettes that illustrate how the Data Model can be used by educators, vendors, and researchers. In addition to the vignettes below, the Data Model website contains "use case scenarios" that explain in more detail the steps a user could take in applying the Data Model to a specific usage scenario.

Educator Vignette

Challenge

- An LEA wants to evaluate a proposed education software system to see if it will contain the range of data needed to perform its stated function.

- The LEA also wants to ensure that data definitions in the proposed software are consistent with definitions in other software systems already used by the LEA and in reports the LEA is required to generate.

Solution

1. **Use the Data Model to identify requirements.**
    - An LEA staff person goes to the Data Model Search tool that is part of the Data Model website and runs a query. For example, if the proposed system is a special education system, the query term might be "program." A query for the term "program" limited to entity names and class names containing that term would produce the following results:

National Forum on Education Statistics

Publications & Products | PK-12 Data Model | Staff

Home | How to Use Model | Browse Data Model | Search | Model Concepts | Conceptual Modeling | About

- Top Level
  - Association
  - EducationLeadershipArtifact
  - Event
  - InformationExchange
  - InstructionArtifact
  - OperationsArtifact
  - Organization
  - Person
  - Place
  - Program
  - x_Extensions

Search Results For:

entity_name, class_name contains program

Classes

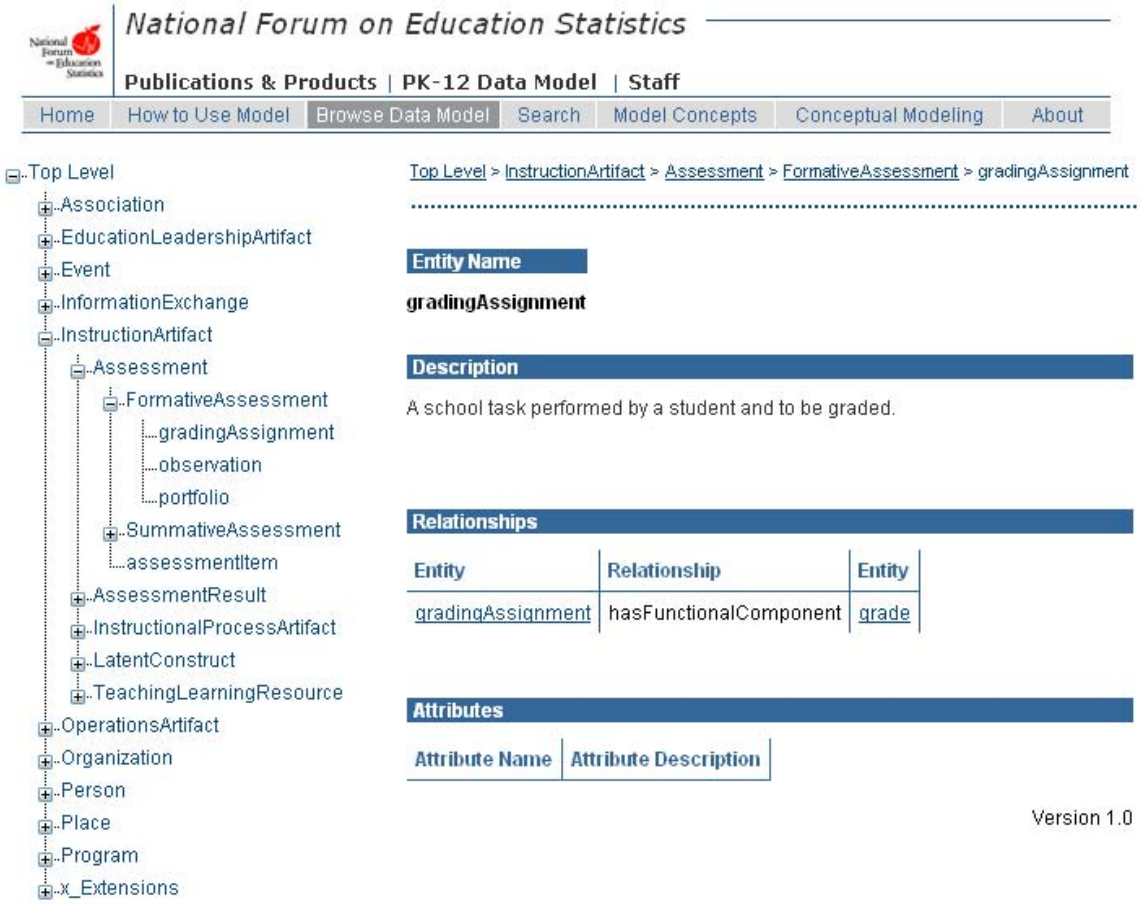
AcademicProgram
AdjunctProgram
ProfessionalDevelopmentProgram
Program
ProgramAssociation
StudentProgram

Entities

alternativeProgram
coCurricularProgram
coCurricularProgram
communityOutreachProgram
educationalProgram
eligibleForProgram
extraCurricularProgram
federalProgram
interventionProgram
ongoingEducationProgram
partnershipProgram
programReport
programService
publicServiceProgram
servedByProgram
supplementalProgram
teacherIncentiveProgram

Version 1.0

- The staff person also uses the navigation bar to find entities of interest, for example "gradingAssignment."



- The staff person follows the relationship links listed for each entity to find other entities that the staff member judges to be related to the requirements of the software system.
- From all of the entities found, the staff person compiles a shorter list of entities that are directly related to the software system. This list is the "Information Requirements" for the system.

2. Perform a gap analysis.

- The staff member, or the vendor, maps the information in the proposed system to the Data Model Information Requirements.
- The staff member compares the Information Requirements to the proposed software system. This is done by using the entity descriptions and attributes for each entity in the Data Model Information Requirements.
- The staff member summarizes the results in a gap analysis document.

3. **Use the Data Model as a canonical conceptual model.**
    - Since the LEA has adopted the Data Model as a standard, all software systems have been mapped to the Data Model; thus, the Data Model is a canonical or standard conceptual model that all physical data models (software designs) refer to.
    - An LEA staff member maps the proposed software system to the Data Model.
    - During the mapping process, the staff member discovers whether the proposed system data definitions are consistent with other software systems used by the LEA. This is possible because the Data Model data definitions represent the definitions used in the LEA.

Vendor Vignette

Challenge

- A vendor wants to ensure that a new education software system that the vendor is building has all the "Information Requirements" needed to fulfill the purpose of the software.

Solution

1. **Use the Data Model to identify requirements.**
    - Using the content domain and the education processes addressed by the new software system, the vendor develops a set of search terms for the Data Model.
    - The vendor goes to the Data Model Search tool that is part of the Data Model website and runs a series of queries.
    - The vendor compiles a list of entities that are relevant to the functions of the new software.
    - The vendor follows the relationship links listed for each entity to find other entities that the vendor judges to be related to the functions of the new software system.
    - The vendor compiles a complete list of entities that are related to the software system. This list is the "Information Requirements" for the system.

2. **Extract a software-specific conceptual model.**
    - Using the entity descriptions, relationships, and attributes for each entity in the Data Model Information Requirements, the vendor builds a conceptual model specific to the new software system.

3. **Build logical and physical models.**
    - Using the software-specific conceptual model, the vendor builds a logical model for the software.
    - Using the logical model, the vendor builds a physical model for the software.

- A gap analysis is performed between the software-specific conceptual model and the physical model to ensure that the physical model has addressed all Information Requirements.

Researcher Vignette

Challenge

- A researcher or a state department program director is creating a research project that will use data available in a state- or district-wide data system. The researcher wants to know what data elements exist now and are expected to exist in the future.

Solution

1. Use the Data Model to identify requirements.

- Using the content domain and the theoretical framework of the research project, the researcher develops a set of search terms for the Data Model.
- The researcher goes to the Data Model Search tool that is part of the Data Model website and runs a series of queries.
- The researcher compiles a list of entities that are relevant to the research.
- The researcher follows the relationship links listed for each entity to find other entities that are related to the research project.
- The researcher compiles a complete list of entities that are related to the research. This list is the "Information Requirements" for the research project.
- Attributes for each entity that are relevant to the research project are also listed.

2. Prepare a research design.

- The researcher can now use the Information Requirements to prepare a proposed research design.
- The Information Requirements in the research design are investigated for availability, or a plan for collecting the data is developed.

Appendix A: Q&A

1. Are the naming conventions used in this data model generally accepted in fields outside of education?
Naming of classes, entities, and attributes in the Data Model was influenced by the well-documented and accepted ISO/EEC 11179 standards. For example, the Education Data Model hierarchy has the following properties:

- Unique names
- Many items that follow the multi-part naming convention, i.e., [Object] [Qualifier] Property RepresentationTerm
- A classification framework in which items can be located using dot notation, such as classname.classname.entity.attribute

Also, class names begin with a capital letter and contain no spaces; entity and relationship names begin with a lower-case letter, followed by camel case, with no spaces; attribute names contain spaces and are capitalized. For example, Person and Client are classes, libraryPatron is an entity, and Library Patron Name is an attribute.

2. Couldn't some items be an entity and attribute at the same time?
Yes. For example, a person listed in a discipline incident could have a victim attribute or a perpetrator attribute; or, we could model the role of victim as an entity. In this case, this is essentially the relationship between a discipline incident and a victim. Attributes of this entity might include injuries sustained and the victim's description of the incident. The Data Model strives to reflect consistent decisions with respect to such issues.

3. Why not just use entity relationship (ER) diagrams or a UML object diagram?
These types of diagrams are used to represent logical or physical models. The Education Data Model is a conceptual map (ontology) with the addition of attribute detail that cannot be represented using ER diagrams or UML object diagrams.

4. How specific are the entities? When do we stop creating classes based on different types and create an entity?
In creating the taxonomy, a number of questions arose concerning the scope and depth of the taxonomy. For example, where do we draw the line between classes and entities? Is "science teacher" an entity in the class of teachers? Or is teacher an entity and the "Subject Taught" a characteristic (attribute) of teachers? Is a "warehouse" a type of building (that is, a coded value in a code set for building type)? Or is a warehouse a separate entity in a class of buildings?

A subjective criterion for creating a lower level entity might be that characteristics of the lower level item must be significantly different from the parent item. Another view might be that characteristics of the child item must be a significant addition to the parent characteristics. So teacher is a type of person and is a child of the class of persons because the characteristics for teachers are a significant addition to the person attributes.

In order to clarify the distinction between a class and an entity, the following guidelines were used in developing the Data Model:

Entity

- An entity may represent a table in an ER design.
- An entity may represent a class (not an instance) in an object-oriented (OO) design. The word "class" has a different meaning in OO design than in the Data Model.
- An entity should fit as part of the collection of entities that make up a class.
- The entities in a class do not make up the complete set of possible entities for the class. In reality, there are an infinite number of possible entities for each class.

Class

- Classes represent concepts.
- Sub-classes make up the complete set within their class. For example, Area and Location make up the complete set of Places in the Data Model.

As an example, there are two teacher-type entities located within the taxonomy in the following way:

Person → Provider → Teacher
Person → Provider → Substitute Teacher

Then shouldn't we also have:

Person → Provider → Teacher → High School Teacher?

or,

Person → Provider → Certified Teacher?

One guideline to use in deciding whether to create the next level of detail, or whether to create a variation of an entity, involves anticipating when a different table for the concepts would be used in a physical or logical model.

Is it reasonable to think that teachers and substitute teachers would be tracked in two separate tables? The answer, in this case, is yes because substitute teachers are different enough from regular teachers that they are typically tracked by different systems or by different rules.

On the other hand, high school teachers and middle school teachers would best be tracked in a single table with a coded field that indicates the level of their teaching assignment. Similarly, certified teachers could be tracked as a status field in a teacher table or as a link to a certifications table.

Appendix B: Glossary

Attribute
Information about an entity that can be measured, classified, or described.

Class
A set, collection, group, or configuration containing members regarded as having certain attributes or traits in common; a kind or category.

Concept map
Describes the structure of the Data Model. A concept map represents a logical set of relationships among classes, sub-classes, and entities.

Conceptual data model
A model that focuses on the meaning of concepts and the relationships among concepts important to address user needs. A conceptual model is not constrained by any kind of database technology.

Entity
The constructs that need to be tracked, measured, and described by software systems in order support education processes.

Logical data model
A logical data model is a representation of the concepts in the information domain but takes into account the technology that will be used. In most cases today, the technology being used to represent information is a relational database; therefore, logical models often have tables, columns, and foreign key relationships.

OWL
Web Ontology Language (OWL) is an XML language for expressing relationships and is a standard published by the W3C web standards group. The OWL language represents both a taxonomy of entities, as well as multiple types of simultaneous relations among the entities.

Physical data model
A physical data model implements the conceptual data model using a particular technology, such as a relational database. The physical model also takes into account local requirements for the system as well as conventions and requirements of the database software.

Sub-class
A sub-class is a class that is a subcategory of another class; for example, primate is a sub-class of mammal. The sub-class itself can also have sub-classes; for example, chordate is a sub-class of the class primate. In the Data Model, a Staff Member is a sub-class of Service Provider and Service Provider is a sub-class of Person.

Bibliography

Boudett, Kathryn Parker and Steele, Jennifer L., eds. 2007. *Data Wise in Action: Stories of Schools Using Data to Improve Teaching and Learning*. Cambridge, MA: Harvard Education Press.

Daconta, Michael C., Orbst, Leo J., and Smith, Kevin T. 2003. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Hoboken, NJ: Wiley Press.

Maguire, Joe 2008. *Data Modeling: A Necessary and Rewarding Aspect of Data Management*. Midvale, UT: Burton Group.

Morville, Peter 2005. *Ambient Findability*. Sebastopol, CA: O'Reilly Media, Inc.

National Forum on Education Statistics website: http://nces.ed.gov/forum/.

National Center for Education Statistics (NCES) Handbooks Online: http://nces.ed.gov/programs/handbook/.

Schools Interoperability Framework Association (SIFA): http://www.sifinfo.org/

Stewart, Darin L. 2008. *Building Enterprise Taxonomies*. Portland, OR: Mokita Press.

Suggested Upper Merged Ontology (SUMO). http://www.ontologyportal.org/.

Universal Data Element Framework (UDEF): The Open Group. http://www.opengroup.org/udef/.